Compressive Sensing with Side Information: Analysis, Measurements Design, and Applications

Meng-Yang Chen

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Electronic and Electrical Engineering

University College London

February 6, 2018

I, Meng-Yang Chen, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Compressive sensing is a breakthrough technology in view of the fact that it enables the acquisition and reconstruction of certain signals with a number of measurements much lower than those dictated by the Shannon-Nyquist paradigm. It has also been recognised in the last few years that it is possible to improve compressive sensing systems by leveraging additional knowledge – so-called side information – that may be available about the signal of interest.

The goal of this thesis is to investigate how to improve the acquisition and reconstruction process in compressive sensing systems in the presence of side information. In particular, by assuming that both the signal of interest and the side information obey a joint Gaussian mixture model (GMM), the thesis focuses on the analysis and the design of linear measurements for two different scenarios: i) the scenario where one wishes to design a linear projection matrix to capture the signal of interest; and ii) the scenario where one wishes to design a linear projection matrix to capture the side information. In both cases, we derive sufficient and (occasionally) necessary conditions on the number of measurements needed for the reliable reconstruction in the low-noise regime and we also derive linear measurement designs that are close to optimal.

Numerical results are presented with synthetic data from both Gaussian and GMM distributions and with real world imaging data that confirm that analysis is well aligned with practice. We also showcase our measurement design scheme can lead to significant improvement on the application example associated with the reconstruction of high-resolution RGB images from gray scale images using low-resolution, compressive, hyperspectral measurements as side information.

Impact Statement

This thesis focuses on the analysis and design of new compressive sensing schemes where one wishes to sense and acquire a target signal but one also has access to other additional sources of information related to the target signal. These problems arise in various domains such as in medical imaging, earth imaging, or even in art investigation.

This thesis concentrates on the design of linear measurements for two different scenarios: i) the scenario where one wishes to design a linear projection matrix associated to the signal of interest; and ii) the scenario where one wishes to design a linear projection matrix associated to the side information. It also shows that designed measurements can lead to considerable performance gains over the traditional random measurement strategies.

We believe that our proposed linear measurement designs can have impact on the performance of various real-world applications beyond the remote sensing ones illustrated in this thesis. Of particular relevance, it can lead to better image reconstruction quality in multi-modal medical imaging applications where one wishes to acquire various modalities such as MRI and PET images.

Acknowledgements

I am very grateful to many people that have contributed to this thesis. My deepest gratitude goes first and foremost to my supervisor, Dr. Miguel Rodrigues, for giving me the opportunity of pursuing my career goal in his research team. I would like to thank him for his constant encouragement, guidance, helpful comments and support throughout my entire PhD studies.

In addition, I would like to acknowledge the opportunity of joining the team of Communications and Information Systems Group (CISG) in the Department of Electronic and Electrical Engineering, at University College London (UCL). I would also like to give my sincere gratitude to the support of Dr. Francesco Renna, and Dr. João Mota, for their insightful suggestions and valuable comments, that were of paramount importance to my research. In addition, I would like to express my indebtedness to my colleagues, Jure, Pingfan, Zahra, and Jaweria, for all the support of my PhD training and sharing the lab with me.

Moreover, many thanks are also to the thesis review committee, for their valuable comments with their enthusiasm and patience on this thesis.

At last, I would like to thank all my family and friends for the caring and support over the years. Most importantly of all, my deepest thanks would go to my beloved family, for giving me unconditional encouragement and providing infinite love and support throughout my PhD studies. In particular, my wife Wei-Ting Cho and my mother Feng-Ying Yeh have been so supportive during the period of this long journey. Without them, I would not have been able to complete this thesis.

Contents

List of Figures

List of Tables

Notation

Boldface upper-case letters ($\mathbf{X}$)	Matrices
Boldface lower-case letters ($\mathbf{x}$)	Column vectors
$\mathbf{I}_n$	Identity matrix of dimension $n \times n$
$\mathbf{0}_{m \times n}$	All-zero-entries matrix of dimension $m \times n$
$\\|\mathbf{x}\\|_0$	$\|supp(\mathbf{x})\|$
$\\|\mathbf{x}\\|_p$	p-norm of vector $\mathbf{x}, 1 < p < \infty$
$\mathcal{O}$	Big $\mathcal{O}$ notation where $g(\mathbf{x}) = \mathcal{O}(f(\mathbf{x}))$ if $\lim_{\mathbf{x} \to \infty} \frac{g(\mathbf{x})}{f(\mathbf{x})} = c$, where c is a constant
o	Small o notation where $g(\mathbf{x}) = o(f(\mathbf{x}))$ if $\lim_{\mathbf{x} \to \infty} \frac{g(\mathbf{x})}{f(\mathbf{x})} = 0$
$< f, g >$	Inner product
$\log(\cdot)$	Natural logarithm
min	Minimum operator
max	Maximum operator
$\mu(\mathbf{\Phi})$	coherence of a given matrix $\mathbf{\Phi}$
spark($\mathbf{\Phi}$)	spark of a given matrix $\mathbf{\Phi}$
$\mathcal{T}(\cdot)$	hard-thresholding operator
$\| \cdot \|$	Cardinality of a set
$(\cdot)^{-1}$	Inverse operator
$(\cdot)^{\mathrm{T}}$	Transpose operator
$(\cdot)^{\dagger}$	Moore-Penrose pseude-inverse
$\mathrm{rank}(\cdot)$	Rank
$\mathbb{E}[\cdot]$	Expectation operators
$\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$	The Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$

$\mathrm{Im}(\cdot)$	The (column) image of a matrix
$\mathrm{Null}(\cdot)$	Null space of a matrix
$\dim(\cdot)$	Dimension of a linear space
$\|\| \cdot \|\|_F$	Frobenius norm of a matrix
$\mathrm{diag}(a_1, \cdots, a_n)$	$n \times n$ diagonal matrix

Abbreviation

GMM	Gaussian Mixture Model
CS	Compressive Sensing
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
CASSI	Coded Aperture Snapshot Spectral Imaging
MMSE	Minimum Mean-squared Error
RIP	Restricted Isometry Property
NP	Non-deterministic Polynomial-time
BP	Basis Pursuit
BPDN	Basis Pursuit Denoising
LASSO	Least Absolute Shrinkage and Selection Operator
AWGN	Additive White Gaussian Noise
MP	Matching Pursuit
OMP	Orthogonal Matching Pursuit
SOMP	Stagewise Orthogonal Matching Pursuit
ROMP	Regularized Version of Orthogonal Matching Pursuit
TBOMP	Tree-based Orthogonal Matching Pursuit
CoSaMP	Compressive Sampling Matching Pursuit
HHS	Heavy Hitters on Steroids
IHT	Iterative Hard Thresholding
PDF	Probability Density Function
PMF	Probability Mass Function
JSM	Joint Sparsity Model
DCS	Distributed Compressive Sensing

MSE	Mean-squared Error
MAP	Maximum a Posteriori Estimation
EM	Expectation-maximization
PSNR	Peak Signal-to-noise Ratio
GSVD	Generalized Singular Value Decomposition
SNR	Signal-to-noise Ratio
FPA	Focal Plane Array
LISS	Linear Imaging Self Scanning
WiFS	Wide Field Sensor
IRS	Indian Remote Sensing
HRV	High Resolution Visible
CASSI	Coded Aperture Snapshot Spectral Imaging
CDL	Coupled Dictionary Learning

Chapter 1

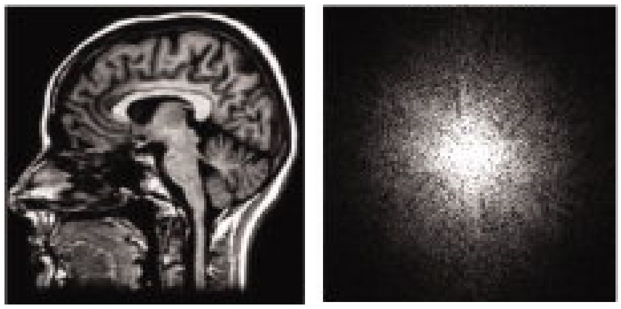
Introduction

1.1 Motivation

Modern signal processing systems – which manipulate digital versions of real-world analog signals – rely on analog-to-digital and digital-to-analog converters to convert the signal from the analog to the digital domain and vice-versa. This acquisition process often involves two stages: First, one converts the original typically real-valued continuous-time signal onto a discrete-time one via a periodic sampling process. Second, one converts the real-valued discrete-time signal onto a quantized discrete-time one via a quantization process [1].

The Shannon-Nyquist paradigm has laid the foundations for the signal sampling process. It asserts that one can perfectly reconstruct a continuous-time signal from the sampled discrete-time one provided that the sampling rate is at least twice the Fourier bandwidth of the original signal [2]. Other variations of this paradigm, e.g. applicable to bandpass signals, have also been established in the past [3].

However, more recently the compressive sensing paradigm has demonstrated that it is often possible to capture a signal at rates much lower than the Shannon-Nyquist rate provided that the signal admits some structure, e.g. a sparse representation in some basis or frame [4, 5]. The acquisition process often involves a linear projection operation and the reconstruction process involves the resolution of a convex optimization problem [6]. Since many natural and man-made signals indeed obey such structures, e.g., natural images obey a sparse representation in a wavelet basis, magnetic resonance imaging (MRI) data are naturally sparse in the Fourier domain as depicted in Fig. 1.1 [7], and the radar target reflectivity through radar process is often sparse in the Fourier or wavelet domain [8], has revled to a surge of interest in the theory and practice of compressive sensing that has spanned over a decade. Compressive sensing now has applications in various areas such as image compression [9], medical imaging [7, 10, 11] and remote sensing [8, 12, 13].



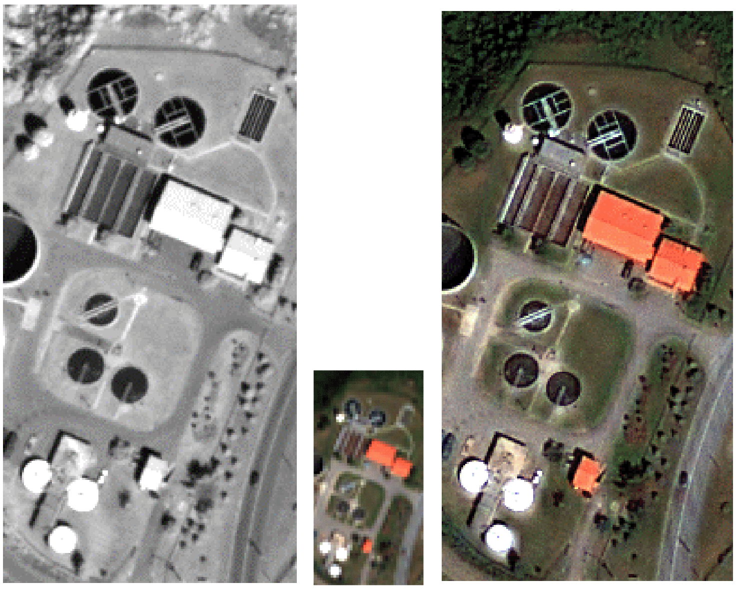
(a) MRI data (b) MRI data in Fourier domain

Figure 1.1: Original MRI (a) and MRI data in Fourier domain (b) [7].

(a) PET scan. (b) MRI scan. (c) PET-MRI scan.

Figure 1.2: PET, MRI, and PET-MRI scans of a small animal [14].

Yet, there are various application scenarios where one wishes to acquire and reconstruct some target signal but one also has access to other sources of information, i.e. other signals that are correlated to the target signal. For example, in the medical application scenarios, prior MRI scans can be used as the side information in the reconstruction of current MRI scans [11]. References [14, 18, 19] report that the PET-MRI scanners can simultaneously sense and acquire functional positron emission tomography (PET) data accompanied with anatomical



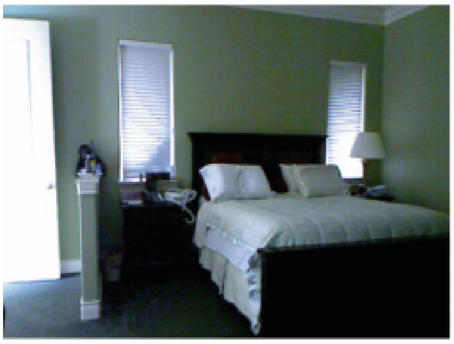
(a) Panchromatic Image. (b) Multispectral Image. (c) Pan-sharpening result

Figure 1.3: QuickBird images and pan-sharpening results [15].

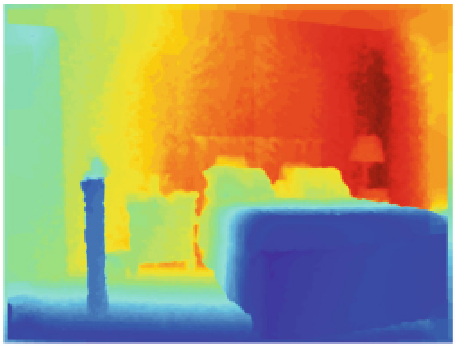
(a) Multispectral image of wavelength 440 nm. (b) RGB image. (c) Reconstruction result with side information. (d) Reconstruction result with no side information.

Figure 1.4: Input multispectral (a), RGB images (b), and the corresponding reconstruction results (c) and (d) [16].

or functional MRI data in the same imaging session where the integrated data not only provides anatomic information but also measures a range of activity as shown in Fig. 1.2. In the remote sensing domain, pan-sharpening is an image fusion process that involves integrating a high-resolution panchromatic image and a low-resolution multispectral image, in order to sharpen the spatial resolution of a multispectral image [20]. Such pan-sharpening technique is based on the fact that earth information for the same geographical region can be represented by multiple



(a) RGB image.



(b) Depth version image.

Figure 1.5: RGB with depth images [17].

(a) Non flash image.

(b) Flash version image.

Figure 1.6: Non-flash and flash images [17].

heterogeneous images corresponding to different band versions [21, 22]. Fig. 1.3 illustrates a pan-sharpening results of a low-resolution multispectral image having a spatial resolution of 2.4 m with the corresponding panchromatic image having a spatial resolution of 0.6 m [15]. In turn, [16] considers the low-resolution multispectral images as the target signal and the corresponding high-resolution RGB versions of the same subject serve as the side information as depicted in Fig. 1.4. In particular, Fig. 1.4(c) illustrates that their proposed design with side information can significantly improve the reconstruction performance with respect to the case of no side information as shown in Fig. 1.4(d). Moreover, the *guided* or *joint image filtering* is a process that uses an additional guidance signal as a structure prior and transfers that structure to an input image, in order to preserve edges and boundaries [23, 17]. For a target image, the guidance image can be the input data itself or other images from different modalities, e.g., RGB with depth images

(Fig. 1.5), flash with non-flash images (Fig. 1.6) [17].

The presence of additional sources of information leads naturally to the questions:

- *How can we leverage such additional sources of information in order to improve the signal acquisition and reconstruction process?*

This thesis attempts to answer these questions by investigating new linear projection designs for compressive sensing systems in the presence of side information.

1.2 Organization of the Thesis

The remainder of this thesis is organized as follows.

In Chapter 2, we provide an overview of major advances associated with compressive sensing. In particular, we introduce the sensing model, signal reconstruction algorithms, and signal reconstruction guarantees both in scenarios where one does not have access to side information and scenarios where one does have access to additional sources of information. We also introduce compressive sensing with structured signal models where one has additional structural knowledge about the signal beyond sparsity. Of particular relevance, we present recent advances in compressive sensing – both with and without side information – in scenarios where the signals obey a GMM model.

The main contributions of the thesis are reported in Chapters 3, 4 and 5. In Chapter 3 we consider the design of projection kernels for scenarios where one wishes to capture and reconstruct a target signal in the presence of another signal – the side information – that is correlated with the target signal. In particular, this chapter focuses on projection kernel designs to capture the target signal when (1) the decoder or (2) both the decoder and the encoder have access to another signal – the side information – that is correlated with the target signal. We provide a number of necessary and sufficient conditions for the reconstruction error to approach zero in the low-noise regime when the target signal and the side information signal obey a joint GMM model. We also provide a range of numerical results – both with synthetic and real data – that showcase that our analysis aligns with simulation. The Appendix outlines the proofs of the main theorems appearing in Chapter 3.

In Chapter 4 we also consider the design of projection kernels for scenarios where one wishes to capture and reconstruct a target signal in the presence of another signal – the side information with a focus on kernel designs to capture the side information signal. We also provide a number of necessary and sufficient conditions for the reconstruction error to approach zero in the low-noise regime when the target signal and the side information signal obey a joint GMM model as well as a range of numerical results – both with synthetic and real data – that showcase that our analysis aligns with simulation. Of particular relevance, this chapter demonstrates that it is possible to obtain substantial improvements in reconstruction performance by adopting optimal kernel designs in comparison to random ones. The Appendix also outlines the proofs of the main theorems appearing in Chapter 4.

Chapter 5 concentrates on the application of our kernel designs in a real-world application that involves the reconstruction of a high-resolution RGB image from a gray-scale one in the presence of low-resolution, compressive, hyperspectral measurements as side information. In particular, we illustrate how to apply the results of previous chapters to design compressive hyperspectral measurements using the state-of-the-art Coded Aperture Snapshot Spectral Imaging (CASSI) architecture. We also illustrate that our CASSI designed measurements can lead to substantial gains over standard random CASSI measurements. These results showcase the potential of our proposed design approach to considerably improve the performance of compressive sensing systems.

Finally, Chapter 6 summarizes the main contributions of this work as well as directions for future research.

1.3 Contributions of the Thesis

This thesis studies how to design measurements for compressive sensing systems where one wishes to sense and reconstruct a target signal in the presence of side information. In particular, the thesis focuses on scenarios where both the target signal and the side information signal obey state-of-the-art structured signal model – a joint GMM model – in *lieu* of standard sparsity models.

The main contributions of this work are:

1. We study compressive sensing with side information systems where one wishes to capture and reconstruct a target signal in the presence of a side information signal. Here, we consider both scenarios where one wishes to optimally capture the target signal and scenarios where one wishes to optimally capture the side information instead.

2. In the scenario where one wishes to design a linear projection kernel to optimally capture the target signal, we provide sharp sufficient and necessary conditions on the number of measurements needed for the reconstruction minimum mean-squared error (MMSE) to approach zero both in the case where the decoder or both the encoder and the decoder have access to side information.

3. In the scenario where one wishes to design a linear projection kernel to optimally capture the side information signal, we also provide sufficient and (occasionally) necessary conditions on the number of measurements needed for the reconstruction minimum mean-squared error (MMSE) to approach zero in the case where the decoder has access to side information. We also provide closed-form linear projection designs that exhibits substantial performance in comparison to standard random ones.

4. We showcase a number of simulation results both with synthetic data and real data (imagery data) in order to showcase that our analytic results are very well aligned with numerical ones. These simulation results also showcase that it is possible to obtain substantial reconstruction error performance improvements by optimally capturing the side information signal.

5. Finally, we concentrate on a real-world pan-sharpening use-case arising in remote sensing applications that involves the reconstruction of a RGB image from a gray-scale one in the presence of hyperspectral data. In particular, we showcase how to use our proposed framework and results to design compressive measurements for the CASSI architecture. We also showcase that such designed measurements lead to substantial reconstruction improvements in comparison to random ones.

These contributions have led to the following publications:

- M-Y. Chen, F. Renna, and M. R. D. Rodrigues, "Compressive Sensing with Side Information: How to Optimally Capture this Extra Information?", *IEEE Transactions on Signal Processing (TSP), 2017 (second round revision).*

- M-Y. Chen, F. Renna, and M. R. D. Rodrigues, "On the Design of Linear Projections for Compressive Sensing with Side Information", has accepted to the *IEEE International Symposium on Information Theory (ISIT), July 2016.*

- M-Y. Chen, F. Renna, and M. R. D. Rodrigues, "Signal Reconstruction in the Presence of Side Information: The Impact of Projection Kernel Design", has accepted to the *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Mar. 2016.*

Chapter 2

Compressive Sensing Systems

This chapter provides an overview of compressive sensing systems. In particular, it overviews the major advances in sampling theory that eventually led to compressive sensing. It also describes basic compressive sensing systems, including the signal sensing and reconstruction process, as well as compressive sensing systems that leverage the presence of side information.

Finally, this chapter reviews compressive sensing systems that leverage other structured signal models beyond signal sparsity.

2.1 Introduction

The pioneering work done by H. Nyquist, C.E. Shannon, V.A. Kotelńikov, E.T. Whittaker and J.M. Whittaker on sampling continuous-time band-limited signals led to the theoretical foundation of the digital revolution [24, 2, 25, 26]. These works define how a numerical sequence of samples can capture all the information from a continuous-time band-limited signal by using an appropriate sampling rate. The most commonly used sampling theorem, known as Nyquist–Shannon sampling theorem, states that it is possible to accurately recover the band-limited continuous-time signal from the discrete-time one provided that the sampling rate is twice the Fourier bandwidth of the signal of interest [2].

In particular, this sampling theorem – which was first introduced by C. E. Shannon – states that:

Theorem 1. *[2] If a function $f(t)$ contains no frequencies higher than B Hertz, then it is completely determined by giving its ordinates at a series of points spaced $(1/2B)$ seconds apart.*

The development of sampling theorem was also implied by the work of Nyquist [24]. Nyquist had proved that a band-limited channel of bandwidth B Hertz can transmit at limiting pulse rate of 2B pulses per second, i.e.,

$$f_b \leq 2B, \tag{2.1}$$

where f_b is the pulse frequency (in pulses per second) and B is the bandwidth (in Hertz). However, Nyquist did not explicitly consider the sampling and reconstruction problems. The quantity 2B later often came to be called the *Nyquist rate*.

The sampling theorem has been attributed in the engineering community to numerous authors. The Whittaker–Shannon interpolation formula dated back to the work of E. T. Whittaker in his study of the cardinal functions [26]. In 1933, V.A. Kotelńikov led the formulation and proof of the sampling theorem [25]. In 1935, J. M. Whittaker refined E. T. Whittaker's work, discussing the relation between the cardinal functions and the finite-limit Fourier integral [27]. Hence, the results of his work were very closely associated with the sampling theorem.

Therefore, Nyquist–Shannon sampling theorem, is also known as Nyquist–Shannon–Kotelńikov, Whittaker–Shannon–Kotelńikov, Whittaker–Nyquist–Kotelńikov–Shannon, sampling theorem, due to the fact that this theorem was also discovered independently by E. T. Whittaker, by V.A. Kotelńikov, and by others.

The ability to represent a continuous-time signal via a discrete-time one – whose amplitudes can be further quantized to enable a digital representation – has led to a shift of attention from the analog to the digital domain, thus enabling the creation of sensing systems that are more flexible and robust (e.g. to noise) in relation to their analog counterpart. In particular, it is typical now to represent digitally text, sound, images and video for efficient data processing, transmission, and storage. However, the resulting Nyquist rate can be too costly to be implemented due to fact that it is not possible to build devices for sensing at such required sampling rate in many real-world applications [28], such as imaging, medical imaging, video streaming, remote surveillance, genomic data analysis and meteorology data analysis.

More recently, the so-called Nyquist paradigm has been transcended by a breakthrough sensing modality, known as *compressive sensing* (CS) [4, 5, 6, 29, 30, 31], that offers the means to simultaneously perform data acquisition and reconstruction. Around 2006, E. Candès, T. Tao, J. Romberg and D. Donoho proved that a finite-dimensional signal can be exactly recovered from a small set of linear measurements subject to appropriate conditions. In particular, this emerging paradigm provides nearly perfect reconstruction without any or with minimal loss of information using far fewer measurements than conventional acquisition schemes provided that the signal is sparse in some orthonormal dictionary or frame [4, 5].

Compressive sensing has been the subject of intense study in the previous decade with numerous contributions focusing on the signal sensing process or the signal reconstruction process using tractable [6, 32, 33, 34] or iterative [35, 36, 37] methods. Compressive sensing has also been studied for scenarios where the signal obeys simple structures (e.g. sparsity representations in some basis or frame) or more complex structures such as wavelet trees or manifolds [38, 39, 40, 41]. Other contributions focus on the application of compressive sensing to MR imaging [42] or radar [43, 44].

The following sections overview the field of compressive sensing as well as its extensions.

2.2 Compressive Sensing

2.2.1 Sensing Model

We consider $\mathbf{y} \in \mathbb{R}^m$ is a set of linear measurements of the signal of interest $\mathbf{x} \in \mathbb{R}^n$, obtained via a projection kernel[1] $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ with $m \ll n$. We can denote the compressive sensing system as follows

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{2.2}$$

[1]Throughout the thesis, we will refer to $\mathbf{\Phi}$ as the sensing matrix, measurement matrix, projection matrix and kernel, interchangeably.

in the absence of noise, or

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}, \tag{2.3}$$

in the presence of noise, $\mathbf{n} \in \mathbb{R}^m$ represents the measurement noise vector.

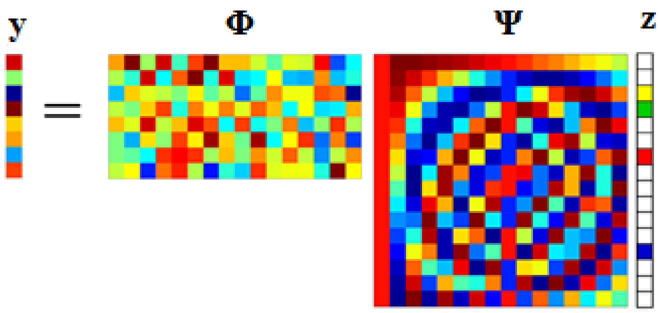


Figure 2.1: Compressive sensing sparse model – I.

It is typically assumed that the signal of interest $\mathbf{x}$ is sparse in some basis or frame. In particular, we write

$$\mathbf{x} = \mathbf{\Psi z}, \tag{2.4}$$

where $\mathbf{z} \in \mathbb{R}^n$ is the sparse representation of the signal $\mathbf{x} \in \mathbb{R}^n$ induced by the basis $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$, i.e. $\|z\|_0 = s \ll n$.

Therefore, we can also express the measurement model in (2.2) as follows

$$\mathbf{y} = \mathbf{\Phi \Psi z} = \mathbf{\Theta z}, \tag{2.5}$$

where $\mathbf{\Theta} = \mathbf{\Phi \Psi} \in \mathbb{R}^{m \times n}$, which corresponds to the product of the sensing matrix and the sensing basis, is also known as an equivalent sensing matrix (see Figs. 2.1 and 2.2).

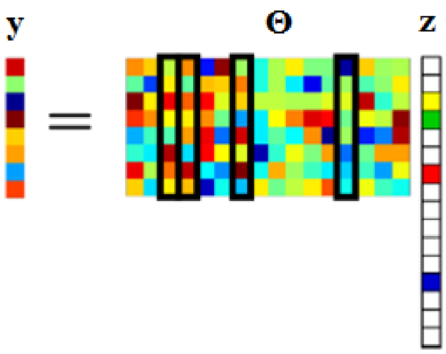


Figure 2.2: Compressive sensing sparse model – II.

In view of the fact that the dimensionality of the linear measurements can be much lower than the dimensionality of the signal of interest, CS asks two fundamental questions:

- How do we construct the measurement matrix $\mathbf{\Phi}$ (or equivalent measurement matrix $\mathbf{\Theta}$) in order to preserve the important information in the signal of interest $\mathbf{x}$ or its sparse representation $\mathbf{z}$?

- How do we recover a high-dimensional signal $\mathbf{x}$ from a small set of linear measurements $\mathbf{y}$?

We start by discussing how to design the measurement matrix to ensure that the information in the signal is well preserved in the measurement space, in order to guarantee successful reconstructions. In particular, we will introduce a number of properties associated with the linear measurement matrix that can dictate whether or not one can reconstruct the signal of interest from the linear measurements correctly[2].

The first property is the *coherence* of the measurement matrix.

Definition 1. *[45, 46] The coherence $\mu(\mathbf{\Phi})$ of a given matrix $\mathbf{\Phi}$ corresponds to the largest absolute inner product between any two columns of $\mathbf{\Phi}$:*

$$\mu(\mathbf{\Phi}) = \max_{1\leq i\neq j\leq n} \frac{| < \phi_i, \phi_j > |}{\|\phi_i\|_2, \|\phi_j\|_2}, \tag{2.6}$$

It is straightforward to show that the coherence of a matrix is in the range $\mu(\mathbf{\Phi}) \in \left[\sqrt{\frac{n-m}{m(n-1)}}, 1\right]$. In particular, when $m \ll n$ the lower bound can be approximated to $1/\sqrt{m}$. The relevance of this notion is due to the fact that the higher the incoherence of a matrix the higher the recovery performance of a reconstruction algorithm [47].

The second property is the *spark* of the measurement matrix.

Definition 2. *[45] The spark$(\mathbf{\Phi})$ of a given matrix corresponds to the smallest number of columns from $\mathbf{\Phi}$ that are linearly dependent.*

[2] We will consider in the rest of the section that the signal of interest is sparse in the canonical basis $\mathbf{\Psi} = \mathbf{I}$.

It has been established that the spark of a matrix is linked to the Kruskal's rank of the matrix, i.e. the Kruskal rank of $\mathbf{\Phi}$ is equal to spark$(\mathbf{\Phi}) - 1$.

The important of this notion is also due to the fact that it can provide reconstruction guarantees in compressive sensing.

Theorem 2. *[45] If spark*$(\mathbf{\Phi}) > 2s$*, then for each measurement vector* $\mathbf{y}$*, there exists at most one signal* $\mathbf{x}$ *with* $\|\mathbf{x}\|_0 \leq s$ *such that* $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$*.*

Note that by connecting the coherence and the spark of a matrix [9], we can also establish a relationship between the mutual coherence and the spark of a matrix.

Lemma 3. *[45] For any matrix* $\mathbf{\Phi}$*,*

$$\text{spark}(\mathbf{\Phi}) \geq 1 + \frac{1}{\mu(\mathbf{\Phi})}. \tag{2.7}$$

Therefore, we can also establish a recovery guarantee for linearly measured s-sparse signals[3].

Theorem 4. *[45] If*

$$s < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{\Phi})}\right). \tag{2.8}$$

then for each measurement vector $\mathbf{y} \in \mathbb{R}^m$*, there exists at most one signal* $\mathbf{x}$ *with* $\|\mathbf{x}\|_0 \leq s$ *such that* $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$*.*

Another important question relates to whether one can recover the signal of interest from the signal linear measurements in the presence of noise. This can be addressed by introducing the restricted isometry property (RIP) that has been established by Emmanuel Candes and Terence Tao [30, 4].

Definition 3. *[30, 4] A matrix* $\mathbf{\Phi}$ *has the* (s, δ_s)*-restricted isometry property* $((s, \delta_s)$*-RIP) provided that:*

$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|\mathbf{x}\|_2^2. \tag{2.9}$$

for any s-sparse vector $\mathbf{x}$ *(i.e. any vector* $\mathbf{x}$ *with* $\|\mathbf{x}\|_0 \leq s$*).*

[3]In particular, we denote that a vector $\mathbf{x}$ is s-sparse if it has only s entries different than 0. The number of non-zero elements of x is denoted as $\|\mathbf{x}\|_0$.

The (s, δ_s)-RIP ensures that all submatrices of $\mathbf{\Phi}$ of size $m \times s$ are nearly an isometry, thus being distance-preserving. In this sense, the measurement matrix $\mathbf{\Phi}$ must preserve the lengths of these particular s-sparse vectors.

Regarding the RIP in conjunction with the spark, the matrix can uniquely identify all s-sparse vectors (i.e. any vector $\mathbf{x}$ with $\|\mathbf{x}\|_0 \leq s$) if the matrix $\mathbf{\Phi}$ has the $(2s, \delta_s)$-RIP with $\delta_s > 0$. Such condition implies all sets of $2s$ columns of $\mathbf{\Phi}$ are linear independent, i.e., spark$(\mathbf{\Phi}) > 2s$, in order to approximately preserve the distance between any of s-sparse vectors [48].

The issue with the RIP is that it is hard to determine whether a certain projection matrix $\mathbf{\Phi}$ satisfies it or not [49].

Fortunately, many random matrices – whose entries are drawn i.i.d from a continuous distribution – have been shown to satisfy the restricted isometry property with high probability. For example, Gaussian, Bernoulli matrices are typically used as sensing matrices in CS due to the fact that they can well construct matrices with i.i.d entries with (s, δ_s)-RIP with high probability in order to satisfy these properties under certain conditions [31], if

$$m = \mathcal{O}(s \log(n/s)). \tag{2.10}$$

Moreover, selecting an i.i.d Gaussian $\mathbf{\Phi}$ as a measurement matrix provides two other useful properties. Such a matrix can be used to sense both signals that are sparse in the canonical basis $\mathbf{\Psi} = \mathbf{I}$ as well as signals that are sparse in some other basis $\mathbf{\Psi}$ because $\mathbf{\Theta} = \mathbf{\Phi\Psi}$ would also satisfy the RIP with high probability provided that $m = \mathcal{O}(s \log(n/s))$.

2.2.2 Reconstruction Algorithms

We now focus on concrete algorithms that can be used to recover the signal of interest from the signal linear measurements.

2.2.2.1 ℓ_0-norm Algorithm

With the assumption that the signal $\mathbf{x}$ is sparse, i.e., $\|\mathbf{x}\|_0 \leq s$, a possible reconstruction procedure to recover the vector $\mathbf{x}$ from noiseless linearly measurements

$\mathbf{y} = \mathbf{\Phi x}$ involves solving the ℓ_0-norm minimization problem [50, 51]:

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{x}\|_0 \\ &\text{subject to } \mathbf{y} = \Phi\mathbf{x}. \end{aligned} \tag{2.11}$$

Unfortunately, this is a very difficult combinatorial optimization problem [35, 52].

Therefore, an alternative approach to recover the vector $\mathbf{x}$ from the linear measurements $\mathbf{y}$ involves replacing the ℓ_0-norm by the ℓ_1-norm thereby converting the computationally intractable optimization problem into a convex and tractable one [36]. Of particular relevance, it can be shown that the ℓ_1 based solution is equivalent to the ℓ_0 based one under some conditions [4, 29, 53].

2.2.2.2 Convex ℓ_1-norm Algorithms

In the absence of noise, we can recover the sparse vector $\mathbf{x}$ from the noiseless linear measurement vector $\mathbf{y}$ by solving the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{x}\|_1 \\ &\text{subject to } \mathbf{y} = \Phi\mathbf{x}. \end{aligned} \tag{2.12}$$

This optimization problem is typically known as *basis pursuit* (BP) [36]. BP is a mathematical optimization process for solving sparse approximation problems of underdetermined system of linear equations in polynomial time.

In contrast, in the presence of noise, we can recover the sparse vector $\mathbf{x}$ from the noisy linear measurement vector $\mathbf{y}$ by solving the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{x}\|_1 \\ &\text{subject to} \|\mathbf{\Phi x} - \mathbf{y}\|_2^2 \leq \epsilon, \end{aligned} \tag{2.13}$$

where ϵ is typically small and proportional to an estimate of the standard deviation of any noise in the measurements. This optimization problem is known as *basis pursuit denoising* (BPDN).

Another option is to consider an unconstrained version of the ℓ_1-optimization, given by

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \quad \lambda > 0, \tag{2.14}$$

where λ is a parameter that controls the trade-off between sparsity and reconstruction fidelity. In particular, it can be shown that this formulation is equivalent to the formulation in (2.13) for some value of λ. We can note that an appropriate choice of the parameter λ can be seen as a coincident of *least absolute shrinkage and selection operator* (LASSO) [32].

Despite the fact that these optimization problems involve using a ℓ_1-norm in *lieu* of a ℓ_0-norm, they can nonetheless yield a very good approximation to the true sparse vector subject to appropriate conditions. These conditions can be expressed either in terms of the coherence of the sensing matrix or the RIP.

Recovery guarantees based on coherence.

We first show reconstruction guarantees associated with the BP algorithm based on coherence.

Theorem 5. *[45] Consider the signal* $\mathbf{x}$ *is a* s*-sparse vector, i.e.* $\|\mathbf{x}\|_0 \leq s$*, and linear measurements* $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$*, it has been established that BP in* (2.12) *yields the solution* $\hat{\mathbf{x}} = \mathbf{x}$ *(i.e. the solution is equal to the original sparse vector) provided that*

$$s < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{\Phi})}\right). \tag{2.15}$$

In the presence of noise, we have the following result.

Theorem 6. *[54] Consider the signal* $\mathbf{x}$ *is a* s*-sparse vector, i.e.* $\|\mathbf{x}\|_0 \leq s$*, and linear measurements* $\mathbf{y} = \mathbf{\Phi}\mathbf{x}+\mathbf{n}$ *where* $\|\mathbf{n}\|_2 \leq \epsilon$*. It can be established that BPDN in* (2.13) *can deliver an estimate* $\hat{\mathbf{x}}$ *of* $\mathbf{x}$ *that satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq \frac{4\epsilon^2}{1 - \mu(\mathbf{\Phi})(4s - 1)}, \tag{2.16}$$

provided that $s < (1 + 1/\mu(\mathbf{\Phi}))/4$*.*

Theorem 7. *[55] Consider the signal* $\mathbf{x}$ *is a* s*-sparse vector, i.e.* $\|\mathbf{x}\|_0 \leq s$*, and linear measurements* $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$ *where* $\|\mathbf{n}\|_2 \leq \epsilon$*. Likewise, it has also been* *established that LASSO in* (2.14) *can deliver an estimate* $\hat{\mathbf{x}}$ *of* $\mathbf{x}$ *that satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_\infty < (3 + \sqrt{\frac{3}{2}})\epsilon \approx 4.22\epsilon, \tag{2.17}$$

provided that $s < 1/3\mu(\Phi)$ *and* $\lambda = 2\epsilon$.

In turn, in the presence of AWGN, we have the following result.

Theorem 8. *[55] Consider the signal* $\mathbf{x}$ *is a* s*-sparse vector, i.e.* $\|\mathbf{x}\|_0 \leq s$, *and linear measurements* $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$ *where* $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma^2)$ *represents AWGN. It has also been established that LASSO in* (2.14) *with probability exceeding* $(1 - \frac{1}{(m-s)^\alpha})(1 - e^{-s/7})$, *it can deliver an estimate* $\hat{\mathbf{x}}$ *of* $\mathbf{x}$ *that satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq (\sqrt{3} + 3\sqrt{2(1+\alpha)\log(m-s)})^2 s\sigma^2, \tag{2.18}$$

provided that $s < 1/3\mu(\Phi)$ *and* $\lambda = \sqrt{8\sigma^2(1+\alpha)\log(m-s)}$ *for some fairly small* $\alpha > 0$.

Recovery guarantees based on RIP.

Recovery guarantees can also be expressed in terms of the RIP.

Theorem 9. *[30, 56] Consider the signal* $\mathbf{x}$ *is a* s*-sparse vector, i.e.* $\|\mathbf{x}\|_0 \leq s$, *and noisy linear measurements* $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$ *where* $\|\mathbf{n}\|_2 \leq \epsilon$. *It can be established that BPDN can deliver an estimate* $\hat{\mathbf{x}}$ *of* $\mathbf{x}$ *that obeys*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq C \cdot \epsilon, \tag{2.19}$$

provided that $\delta_{2s} + \delta_{3s} < 1$, *and for some constant* C.

Convex ℓ_1 based algorithms allow one to recover the original sparse vector from the noiseless or noisy linear measurements with a reasonable computationally complexity. In turn, greedy algorithms – discussed next – are less complex than ℓ_1 based algorithms but greedy approaches require more measurements than ℓ_1 based approaches to recover the true solution.

2.2.2.3 Greedy Algorithms

A family of iterative greedy algorithms has been widely used as an alternative reconstruction approach for sparse signal recovery in the CS. In particular, greedy strategies seek for the best local optimal solution in each iteration with the goal of achieving the optimal holistic solution, i.e. they rely on iterative approximation of the signal of interest $\mathbf{x}$ coefficients, and support by either identifying the elements of the support of $\mathbf{x}$ iteratively, one by one, or by improving the estimate of the sparse signal at each iteration, taking into account the mismatch to the measured data [9]. In short, such algorithms solve the reconstruction problem of obtaining an approximate sparse representation step by step, in an iterative fashion.

The *matching pursuit* (MP) algorithm computes the best nonlinear approximation representation to a signal through a sequence of mono-atomic approximation, i.e., the algorithm aims to find the "best matching" projections of multidimensional data based on an over-complete (redundant) dictionary (see Algorithm 1) [35, 36, 37].

Algorithm 1 MP Algorithm

Input: signal $\mathbf{x} \in \mathbb{R}^n$, measurement vector $\mathbf{y} \in \mathbb{R}^m$
Output: Estimated sparse vector $\hat{\mathbf{x}}$
1: **Initialization** $\hat{\mathbf{x}} = 0$, $\mathbf{r} = \mathbf{y}$, $i = 0$
2: **while** halting criterion false **do**
3: $h^i = \mathbf{\Phi}^T \mathbf{r}^i$; {form residual signal estimate}
4: $\mathbf{x}^{i+1} = \mathbf{x}^i + \mathcal{T}(h^i, 1)$; {update signal estimate, where $\mathcal{T}$ represents the hard-thresholding operator}
5: $\mathbf{r}^{i+1} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}^{i+1}$; {update measurement residual}
6: $i = i + 1$
7: **end while**

where $\mathcal{T}(v, K)$ defined as:

$$[\mathcal{T}(v, K)]_i \begin{cases} v_i, & \text{if } |v_i| \text{ is among the K largest values of } |v|; \\ 0, & \textit{otherwise} \end{cases}$$

and $[\cdot]$ represents the i-th element of a given vector.

The *orthogonal matching pursuit* (OMP) algorithm is developed based on the MP algorithm, which employs the process of orthogonalization to guarantee the orthogonal direction of projection in each iteration (see Algorithm 2) [50]. Other algorithms based on MP and OMP have also been proposed in order to further improve the reconstruction performance. For example, Donoho proposed an extension of OMP, called stagewise orthogonal matching pursuit (StOMP) [57]. Needell proposed a regularized version of orthogonal matching pursuit (ROMP) algorithm [51] and the Compressive Sampling Matching Pursuit (CoSaMP) algorithm [58]. A new tree-based orthogonal matching pursuit (TBOMP) algorithm was introduced by La and Do [59].

Algorithm 2 OMP Algorithm

Input: signal $\mathbf{x} \in \mathbb{R}^n$, measurement vector $\mathbf{y} \in \mathbb{R}^m$
Output: Estimated sparse vector $\hat{\mathbf{x}}$

1: **Initialization** $\hat{\mathbf{x}} = 0$, $\mathbf{r} = \mathbf{y}$, $i = 0$, $\Lambda = \emptyset$
2: **while** halting criterion false **do**
3: $h^i = \mathbf{\Phi}^T \mathbf{r}^i$; {form residual signal estimate}
4: $\Lambda^{i+1} = \Lambda^i \cup \text{supp}(\mathcal{T}(h^i, 1))$; {update support with residual, where $\mathcal{T}$ represents the hard-thresholding operator}
5: $\mathbf{x}^{i+1} = argmin_{z:\text{supp}(z) \subseteq \Lambda^{i+1}} \|\mathbf{y} - \mathbf{\Phi z}\|_2$ {update signal estimate}
6: $\mathbf{r}^{i+1} = \mathbf{y} - \mathbf{\Phi x}^{i+1}$; {update measurement residual}
7: $i = i + 1$
8: **end while**

where $\mathcal{T}(v, K)$ defined as:

$$[\mathcal{T}(v, K)]_i \begin{cases} v_i, & \text{if } |v_i| \text{ is among the K largest values of } |v|; \\ 0, & otherwise \end{cases}$$

and $[\cdot]$ represents the i-th element of a given vector.

Recovery guarantees based on coherence.

We first show reconstruction guarantees associated with the OMP algorithm based on coherence. It has been established that the OMP algorithm can recover the true sparse signal from noise linear measurements provided that the projection matrix satisfies (2.15) [46].

In the presence of noise, we also have the following result.

Theorem 10. *[60] Consider the signal* $\mathbf{x}$ *is* s*-sparse vector, i.e.* $\|\mathbf{x}\|_0 \leq s$*, and noisy linear measurements* $\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}$*. Suppose that* $s \leq (1/\mu(\mathbf{\Phi}) + 1)/4$ *and* $\epsilon \geq \gamma = \|\mathbf{n}\|_2$*. Then, the signal estimate delivered by OMP in Algorithm 2, where* *the halting criterion is* $\mathbf{r} = \|\gamma\|_2$*, obeys*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{\gamma}{\sqrt{\mu(\mathbf{\Phi})(s-1)}}, \tag{2.20}$$

provided that $\gamma \leq A(1-\mu(\mathbf{\Phi})(2s-1))/2$ *where* A *represents a positive lower bound on the magnitude of non-zero elements of* $\mathbf{x}$.

In turn, in the presence of AWGN, we have the following result.

Theorem 11. *[55] Consider the signal* $\mathbf{x}$ *is* s*-sparse vector, i.e.* $\|\mathbf{x}\|_0 \leq s$*, and noisy linear measurements* $\mathbf{y} = \mathbf{\Phi}\mathbf{x}+\mathbf{n}$*, where* $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\cdot\sigma^2)$ *represents AWGN. Suppose that* $s \leq (1/\mu(\mathbf{\Phi})+1)/4$ *and*

$$\max_{1\leq j\leq n} |\mathbf{x}(j)| \geq \frac{2\sigma\sqrt{2(1+\alpha)\log n}}{1-\mu(\mathbf{\Phi})(2s-1)}, \tag{2.21}$$

for some constant $\alpha > 0$*. Then, with probability at least* $1-(n^{\alpha}\sqrt{\pi(1+\alpha)\log n})^{-1}$*, the signal estimate delivered by OMP in Algorithm 2 after* s *iterations, obeys*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq C\sigma\sqrt{\pi(1+\alpha)s\log n}, \tag{2.22}$$

and its support matches the true s*-entries support of* $\mathbf{x}$*, and for some constant* C.

Recovery guarantees based on RIP.

We first show reconstruction guarantees associated with the OMP algorithm based on RIP.

We can establish the following result applicable to the noiseless setting in (2.2).

Theorem 12. *[61] Consider the signal* $\mathbf{x}$ *is a* s*-sparse vector, i.e.* $\|x\|_0 \leq s$*, and noisy linear measurements* $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$*. Then, the OMP algorithm can exactly recover the signal of interest in* s *iterations provided that* $\mathbf{\Phi}$ *has the* $(s+1, \delta_s)$*-RIP with* $\delta_s < \frac{1}{3\sqrt{s}}$.

We can also establish the following result applicable to the noisy setting in (2.3).

Theorem 13. *[62] Consider the signal* $\mathbf{x}$ *is a s-sparse vector, i.e.* $\|x\|_0 \leq s$*, and noisy linear measurements* $\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}$*. Then, the signal estimate delivered by OMP in Algorithm 2 after 30s iterations, obeys*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 = C\|n\|_2, \tag{2.23}$$

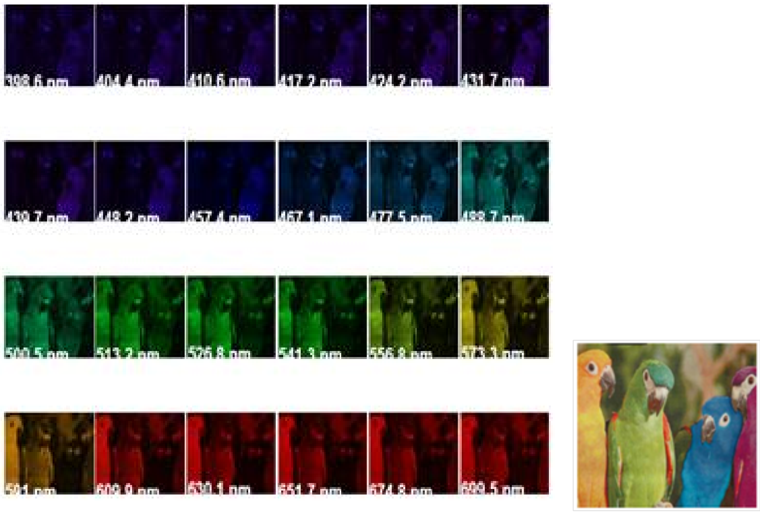
provided that $\mathbf{\Phi}$ *has the* $(31s, \delta_s)$*-RIP, where* $\delta_s < \frac{1}{3}$*, for some constant* C*.*

Other recovery guarantees can also be provided for various other algorithms such as combinatorial approaches (e.g. Heavy Hitters on Steroids (HHS), sub-linear Fourier transform and Chaining Pursuits) and iterative thresholding algorithms (e.g. Iterative Hard Thresholding (IHT), Message Passing algorithm and Belief Propagation). We refer the reader to the following papers [63, 64, 65, 66, 67, 68, 69].

2.3 Compressive Sensing with Side Information

The use of side information – in the form of other signals that may be related to the signal of interest – has also been recently used to improve the performance of CS systems. The rationale has to do with the fact that in various systems one often has access to signals that exhibit some similarity to the signal of interest. For example, in medical imaging one often wishes to reconstruct a certain image modality, e.g. a PET image, but one also has access to another image modality, e.g. an MRI image, associated with the same subject [70]. Moreover, [71] and [72] use information from former scans of the same patient, taken from magnetic resonance images and dynamic tomographic images, respectively, and prove enhanced reconstruction performance with a smaller number of measurements than traditional CS approaches.

In remote sensing, one also often wishes to reconstruct a certain image of a certain scene such as hyperspectral images but one also has access to other images of the same scene such as an RGB image [73] as depicted in Fig. 2.3. Likewise, in art investigation applications one often has access to complementary imaging modalities such as x-ray images, infrared images of art work [74] as depicted in Fig. 2.4. In particular, the separation method is based on unmixing a single X-ray



(a) Hyperspectral images with different wavelengths. (b) RGB image.

Figure 2.3: Hyperspectral images and RGB image.

Figure 2.4: Art-investigation applications.

scan acquired from double-sided paintings, i.e., the (right) X-ray image contains a mixture of components from (left) panels of *Adam and Eve* and (center) the respective paintings.

Recent advances in compressive sensing have proposed different techniques to effectively incorporate side information into the signal sensing or reconstruction [75, 76, 77, 78, 79, 80, 81, 82]. In particular, [75] studies the case when side information is available at the decoder in the form of partial information about the signal support; and [76] studies the case when side information is given as an additional noisy version of the signal of interest. Furthermore, [77, 78] study new reconstruction algorithms – namely, ℓ_1-ℓ_1 and ℓ_1-ℓ_2 reconstruction algorithms – for compressive sensing systems with side information. We elaborate further about the sensing model and these reconstruction algorithms in the sequel.

2.3.1 ℓ_1-ℓ_1 and ℓ_1-ℓ_2 reconstruction algorithms and guarantees

Let us consider a scenario where one has access to a set of linear measurements $\mathbf{y} \in \mathbb{R}^m$ of a sparse signal $\mathbf{x} \in \mathbb{R}^n$ with $m < n$, given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{2.24}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the linear measurement matrix with prior information $\mathbf{w} \in \mathbb{R}^n$ which is similar to the original signal $\mathbf{x}$.

Then, it is possible to reconstruct the signal of interest from the signal measurements by posing the optimization problem

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}}\ \|\mathbf{x}\|_1 + \beta\ g(\mathbf{x} - \mathbf{w}) \\ &\text{subject to}\ \mathbf{y} = \mathbf{A}\mathbf{x}\ , \end{aligned} \tag{2.25}$$

where g represents a function that measures the similarity between original signal $\mathbf{x}$ and the prior information $\mathbf{w}$, in the sense that $g(\mathbf{x}-\mathbf{w})$ is expected to be small provided that $\mathbf{x}$ is similar to $\mathbf{w}$ and β establishes a trade-off between signal sparsity and fidelity to prior information. By considering two specific, convex models for g – $g_1 := \|\cdot\|_1$ and $g_2 := \frac{1}{2}\|\cdot\|_2^2$ where $\|z\|_2$ is the ℓ_2-norm, the optimization problem (2.25) reduces to the optimization problems

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}}\ \|\mathbf{x}\|_1 + \beta\|(\mathbf{x} - \mathbf{w})\|_1 \\ &\text{subject to}\ \mathbf{y} = \mathbf{A}\mathbf{x}\ , \end{aligned} \tag{2.26}$$

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}}\ \|\mathbf{x}\|_1 + \frac{\beta}{2}\|(\mathbf{x} - \mathbf{w})\|_2^2 \\ &\text{subject to}\ \mathbf{y} = \mathbf{A}\mathbf{x}\ , \end{aligned} \tag{2.27}$$

where (2.26) and (2.27) are also referred to as ℓ_1-ℓ_1 and ℓ_1-ℓ_2 minimization problems, respectively.

Reference [77] establishes reconstruction guarantees for the reconstruction algorithms in (2.26) and (2.27).

Theorem 14. *[77] Let* $\mathbf{x} \in \mathbb{R}^n$ *the signal of interest and* $\mathbf{w} \in \mathbb{R}^n$ *be the prior information. Assume* $\bar{h} > 0$ *and that there exists at least one index* i *for which* $\mathbf{x}_i = \mathbf{w}_i = 0$*. Let the entries of* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *be i.i.d. Gaussian with zero mean and variance* $1/m$*. If*

$$m \geq 2\bar{h} \log(\frac{n}{s+\xi}) + \frac{7}{5}(s + \frac{\xi}{2}) + 1, \tag{2.28}$$

then, with probability greater than $1 - exp(-\frac{1}{2}(m - \sqrt{m})^2)$, $\mathbf{x}$ *is the unique solution of* (2.26) *with* $\beta = 1$ *where*

$$\begin{aligned} \bar{h} &:= |\{i \ : \ \mathbf{x}_i > 0, \mathbf{x}_i > \mathbf{w_i}\} \cup \{i \ : \ \mathbf{x}_i < 0, \mathbf{x}_i < \mathbf{w_i}\}|, \\ \xi &:= |\{i \ : \ \mathbf{w_i} \neq \mathbf{x}_i = 0\}| - |\{i \ : \ \mathbf{w_i} = \mathbf{x}_i \neq 0\}|, \end{aligned}$$

and $|\cdot|$ *denotes the cardinality of a set.*

This theorem establishes that – with the availability of side information – ℓ_1-ℓ_1 algorithms can successfully reconstruct the signal of interest with far fewer measurements required by BP algorithms. However, [77] also shows that ℓ_1-ℓ_2 algorithms do not lead to considerable gains.

It is also possible to generalize the results from the setting where the linear measurements, i.e. $\mathbf{y} = \mathbf{A}\mathbf{x}$ are noiseless to the setting where the linear measurements are noisy, i.e. $\mathbf{y} = \mathbf{A}\mathbf{x} + \eta$, by posing the optimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{x}\|_1 + \beta \ g(\mathbf{x} - \mathbf{w}) + \lambda \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2,$$

with $\lambda > 0$ which can be viewed as a Lagrangian relaxation of

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{x}\|_1 + \beta g(\mathbf{x} - \mathbf{w}) \\ &\text{subject to} \ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma, \end{aligned} \tag{2.29}$$

where $\|\eta\|_2 \leq \sigma$.

Other works that study reconstruction algorithms and reconstruction guarantees for compressive sensing with side information appear in [80, 81]. [80, 81] provide sufficient and necessary conditions on the number of random measurements taken from the signal of interest and the side information signal to guarantee that the reconstruction error tends to zero in the low-noise regime. In particular, [81] introduces a GMM model that generalizes the joint sparsity models JSM-1 and JSM-3 proposed for distributed compressive sensing (DCS) in [83] and [84]. Another advance in compressive sensing relates to the use of side information at the encoder [82].

2.4 Compressive Sensing for Structured Signal Models

2.4.1 Overview of Structured Signal Models

The previous sections have overviewed major advances in compressive sensing and compressive sensing with side information applicable to scenarios where the signal of interest is sparse in a canonical basis or else sparse in some basis. However, there are various instances in practice where the signal of interest admits other structures beyond simple sparsity. Recently, the impact of structured signal models has been extensively studied in compressive sensing, which offers the possibility of deriving better compression and reconstruction performance by leveraging additional structure beyond conventional sparsity. Such structured models include wavelet trees [38, 39], manifolds [40, 41], union-of-subspaces [39, 85] or even Bayesian models [86, 76, 69].

The tree modeling is well considered with wavelet dictionaries when the pair of non-zero wavelet coefficients of piecewise smooth signals and images can be naturally organized into a rooted, connected tree structure [87]. The reliable reconstruction for the s-sparse signal can be performed by using a model-based version of the CoSaMP algorithm [58] with required number of projections $\mathcal{O}(s)$ [38].

Manifold-based models have also been shown to reduce the number of measurements needed for the successful recovery. In particular, the number of projections is linearly increasing with the dimension of the manifold $\bar{s}$ and logarithmically with the product of signal size n and parameters that characterize the volume and the regularity [41].

Within the union-of-subspaces, a s-sparse signal is assumed to be modeled as lying in one out of a collection of K linear subspaces with dimension less than or equal to s. Recovery of a signal in a union of subspaces is equivalent to the reconstruction of a block-sparse signal, when the individual subspaces in the union-of-subspaces model are decomposable as the direct sum of a given number of lower dimensional subspaces [85]. In this case, the number of measurements needed for the reliable reconstruction is derived to be $\mathcal{O}(s + \log(2K))$ [39] by mixed ℓ_2/ℓ_1-norm minimization [85].

2.4.2 Reconstruction of GMM Signals with No Side Information

Bayesian inference is a statistical method, which can be used in conjunction with CS for solving the ill-posed data inversion problems associated with the estimation of the underlying signal of interest from the compressed measurements [86, 76, 69]. The term Bayesian CS refers to the case when the recovery algorithm and/or the projection design leverages a statistical description of the signal.

A well-known structured statistical description of a signal is based on the Gaussian Mixture Model (GMM) [86, 88, 89]. There are multiple reasons for adopting a GMM representation, which can be seen as an union of (linear or affine) subspaces, where each subspace is associated with the translation of the image of the (possibly low-rank) covariance matrix of each Gaussian component within the GMM. In fact, a low-rank GMM can also be shown to approximate signals in compact manifolds [40]. Moreover, a GMM model has also been shown to provide state-of-the-art results in practical problems in image processing [90], dictionary learning [40], image classification [91] and video compression [92]. In particular, GMM priors with a moderate number of classes have shown to model reliably real-world data, e.g. patches extracted from images or video frames [92, 93]. Another advantage of GMM models over other structured models relates to the fact that the reconstruction of signals drawn from a GMM distribution can be very effectively performed via an optimal conditional mean estimator available in closed-form [40].

Recent advances in CS have discussed the problem of reconstruction of high-dimensional signals from low-dimensional random linear and noisy measurements, by assuming that the signal of interest is drawn from a GMM [93, 80, 81]. In [93], the proposed CS model determines sharp necessary and sufficient conditions on the minimum number of measurements needed for the MMSE to approach zero in the low-noise regime, where such bounds are tighter and sharper than standard bounds on the number of measurements needed to recover sparse signals. In addition to the case of linear random Gaussian measurements, [93] provides necessary and sufficient conditions for the case of linear kernels that minimize the MMSE. Furthermore, references [80, 81] consider a scenario where a decoder has access both to linear measurements of the signal of interest and to linear features of the side information signal in the CS, assuming both signal of interest and the side information are described by a joint a GMM distribution.

Since this thesis also adopts GMM signal models, in the following, we will review the main contributions in the literature that leverage GMM signal models in the compressive sensing context.

2.4.2.1 System Model

Consider the problem of reconstructing a signal $\mathbf{x} \in \mathbb{R}^n$ from a set of linear, noisy measurements $\mathbf{y} \in \mathbb{R}^\ell$ with $\ell < n$ where

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{w}. \tag{2.30}$$

where $\mathbf{\Phi} \in \mathbb{R}^{\ell \times n}$ is the linear projection kernels and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_\ell \cdot \sigma^2)$ is the additive white Gaussian noise.

In particular, consider that the signal $\mathbf{x} \in \mathbb{R}^n$ obeys a GMM given by

$$\mathbf{x} \sim \sum_{k=1}^{K} p_k \cdot \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(k)}, \mathbf{\Sigma}_{\mathbf{x}}^{(k)}), \tag{2.31}$$

where K is the number of Gaussians in the mixture model, ${\boldsymbol{\mu}_{\mathbf{x}}}^{(k)}$ is the mean of k-th Gaussian, and $\mathbf{\Sigma}_{\mathbf{x}}^{(k)}$ is the covariance of k-th Gaussian, and p_k represents a-priori probability of the k-th mixture component where $\sum_{k=1}^{K} p_k = 1$.

For a Gaussian sources ($K = 1$), we consider that $\mathbf{\Sigma_x}$ represents the possibly low-rank covariance matrix such that $s = \text{rank}(\mathbf{\Sigma_x})$ with $s < n$. In particular, the rank of covariance matrix is given by eigenvalue decomposition of the positive semidefinite covariance matrix

$$\mathbf{\Sigma_x} = \boldsymbol{U}_\mathbf{x}\mathbf{\Lambda_x}\boldsymbol{U}_\mathbf{x}^\dagger = \boldsymbol{U}_\mathbf{x}\text{diag}(\lambda_{\mathbf{x},1}, \cdots, \lambda_{\mathbf{x},s}, 0, \cdots, 0)\boldsymbol{U}_\mathbf{x}^\dagger, \tag{2.32}$$

where the orthogonal matrix $\boldsymbol{U}_\mathbf{x}$ contains the eigenvectors of $\mathbf{\Sigma_x}$, and the diagonal matrix $\mathbf{\Lambda_x} = \text{diag}(\lambda_{\mathbf{x},1}, \cdots, \lambda_{\mathbf{x},s}, 0, \cdots, 0)$ contains the eigenvalues of the covariance matrix, $\lambda_{\mathbf{x},1} \geq \cdots \geq \lambda_{\mathbf{x},s} \geq 0$, so that $s = \text{rank}(\mathbf{\Sigma_x})$ represents the rank of $\mathbf{\Sigma_x}$.

For a GMM sources ($K > 1$), we will also be assuming that $\mathbf{\Sigma}_\mathbf{x}^{(k)}$ represents the possibly low-rank covariance matrix. We consider the linear spaces associated to the images of the covariance matrices $\mathbf{\Sigma}_\mathbf{x}^{(k)}$ for different class labels k such that $s_{\max} = \max_k s_k$ where $s_k = \text{rank}(\mathbf{\Sigma}_\mathbf{x}^{(k)})$ with $s_k < n$.

We are also interested in the reconstruction of the signal of interest $\mathbf{x}$ estimated via the noisy linear measurements $\mathbf{y}$ using the optimal (in MMSE sense) conditional estimator given by

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathbb{E}[\mathbf{x}|\mathbf{y}] = \int_{-\infty}^{+\infty} \mathbf{x}p(\mathbf{x}|\mathbf{y})d\mathbf{x}, \tag{2.33}$$

where $p(\mathbf{x}|\mathbf{y})$ is the *a posteriori* probability density function (pdf) of $\mathbf{x}$ given the measurements $\mathbf{y}$. Therefore, the reconstruction performance is measured in terms of the MMSE which is given by:

$$\text{MMSE}(\sigma^2, \mathbf{\Phi}) = \mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2], \tag{2.34}$$

Note that we express the MMSE explicitly as a function of the noise variances σ^2 and projection kernel $\mathbf{\Phi}$.

We next report reconstruction results both in the scenario where the signal follows a Gaussian distribution ($K = 1$) and in the scenario where the signal follows a GMM distribution ($K > 1$) [93].

2.4.2.2 Gaussian Sources

We now focus on the characterization of the behavior of the reconstruction MMSE in the low-noise regime, i.e., $\sigma^2 \to 0$. The following theorem provides necessary and sufficient conditions on the minimum number of measurements needed for the MMSE to approach zero in the low-noise regime when the source signal is Gaussian.

Theorem 15. *[93] Consider the measurement model in* (2.30)*, where* $\mathbf{x}$ *is described by a Gaussian distribution with mean* $\boldsymbol{\mu}_{\mathbf{x}}$ *and the covariance matrix* $\boldsymbol{\Sigma}_{\mathbf{x}}$*, such that* $s = \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{x}})$*. Assume that the entries of the measurement kernel* $\boldsymbol{\Phi} \in \mathbb{R}^{\ell \times n}$ *are drawn i.i.d. from a zero-mean, fixed-variance, Gaussian distribution.*

1. *When* $\ell \geq s$ *the number of measurements is necessary to guarantee the reconstruction MMSE to approach zero in the low-noise regime.*

2. *When* $\ell \geq s$ *the number of measurements is also sufficient to completely capture the full range of the information source, that guarantees the reconstruction MMSE approaches zero in the low-noise regime.*

Proof: See [93]. ■

It is important to note that, when the n-dimensional signal is known to be s-sparse only, standard CS results show that one provides reliable reconstruction with overwhelming probability with $\mathcal{O}(s \log(n/s))$ for reliable reconstruction; on the other hand, when the signal is known to be Gaussian with a rank-deficient covariance matrix, this result shows that one needs only s measurements for reliable reconstruction. This means that – by leveraging additional knowledge – we can considerably reduce the number of measurements required for reliable reconstruction

2.4.2.3 GMM Sources

We also focus on the characterization of the behavior of the reconstruction MMSE in the low-noise regime. The following theorem now provides necessary and sufficient conditions on the minimum number of measurements needed for the MMSE to approach zero in the low-noise regime when the source is a GMM.

Theorem 16. *[93] Consider the measurement model in* (2.30)*, where* $\mathbf{x}$ *is described by a GMM distribution with the mean* $\boldsymbol{\mu}_{\mathbf{x}}^{(k)}$ *and the possibly low-rank covariance matrix* $\boldsymbol{\Sigma}_{\mathbf{x}}^{(k)}$*, such that* $s_{max} = max_k\ s_k$ *with* $s_k = rank(\boldsymbol{\Sigma}_{\mathbf{x}}^{(k)})$*. We consider random measurement kernel* $\boldsymbol{\Phi} \in \mathbb{R}^{\ell \times n}$ *are taken with i.i.d, zero-mean, Gaussian entries with fixed-variance.*

1. *The necessary condition for the MMSE to approach zero in the low-noise regime, when number of measurements satisfy* $\ell \geq s_{max}$*, i.e.,*

$$\lim_{\sigma^2 \to 0} MMSE^{GM}(\sigma^2) = 0 \Rightarrow \ell \geq s_{max}. \tag{2.35}$$

2. *The sufficient condition on the number of random linear measurements to guarantee the perfect reconstruction in the low-noise regime, then it holds:*

$$\ell > s_{max} \Rightarrow \lim_{\sigma^2 \to 0} MMSE^{GM}(\sigma^2) = 0. \tag{2.36}$$

where $MMSE^{G}(\sigma^2)$ *denotes the MMSE associated with the reconstruction of GMM signals.*

Proof: See [93]. ■

The results suggest that when number of measurements $\ell > s_{\max}$ can lead to MMSE to converge to zero in the low-noise regime. We also observe that when number of measurements $\ell = s_{\max}$ may or may not derive the reconstruction MMSE to approach zero in the low-noise regime, depending on the exact class dependent source covariances.

The results also show that bounds are tighter and sharper than standard CS bounds on the required number of measurements needed for reliable reconstruction. In particular, standard CS results suggest that it is possible to perfectly reconstruct a n-dimensional s-sparse signal with overwhelming probability with $\mathcal{O}(s \log(n/s))$ linear random measurements or projections. In contrast, when the signal is known to be a GMM with a rank-deficient covariance matrix, the results provide a sharp characterization of the region associated to the reconstruction MMSE to approach zero in the low-noise regime, i.e., $\ell > s_{\max}$ is sufficient to derive MMSE to converge to zero as $\sigma^2 \to 0$, and $\ell \geq s_{\max}$ is necessary to derive

MMSE to approach zero as $\sigma^2 \to 0$. This means that – by leveraging the additional knowledge – we can significantly reduce the minimum number of measurements required for reliable reconstruction.

2.4.2.4 Designed Kernels

It is natural to ask whether the projection kernel design can lead to better reconstruction results. The purpose of the design problem attempts to select an appropriate measurement matrix that is able to capture salient signal features in order to enhance reconstruction performance.

Based on the measurement model in (2.30), it is natural to pose the design problem

$$\begin{aligned} &\underset{\mathbf{\Phi}}{\text{minimize}}\ \text{MMSE}(\sigma^2, \mathbf{\Phi}) \\ &\text{subject to}\ \text{tr}(\mathbf{\Phi}\mathbf{\Phi}^{\text{T}}) \leq \ell, \end{aligned} \tag{2.37}$$

where the trace constraint limits the average energy associated to the projection kernel $\mathbf{\Phi}$. We also note that the "optimality" is defined by the optimization problem in (2.37), which follows immediately by leveraging results on the joint optimization of transmitter and receiver for coherent multiple input multiple output (MIMO) [94].

Theorem 17. *[93] Consider the linear measurement model in* (2.30)*, where* $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{\Sigma}_{\mathbf{x}})$*, and* $\boldsymbol{\mu}_{\mathbf{x}}$ *is the mean and* $\mathbf{\Sigma}_{\mathbf{x}}$ *is the covariance matrix,* $s = rank(\mathbf{\Sigma}_{\mathbf{x}})$*. Then, the projection kernel* $\mathbf{\Phi}^{\star} \in \mathbb{R}^{\ell \times n}$ *that solves the optimization problem in* (2.37) *can be expressed as follows:*

$$\mathbf{\Phi}^{\star} = \left[diag(\sqrt{\lambda^{\star}_{\mathbf{\Phi},1}}, \cdots, \sqrt{\lambda^{\star}_{\mathbf{\Phi},\ell}}\ )\mathbf{0}_{\ell\times(n-\ell)} \right] \mathbf{U}_{\mathbf{x}}{}^{\dagger}. \tag{2.38}$$

the squared singular values associated with projection kernel $\mathbf{\Phi}^{\star}$ *are obtained through the water-filling principle [95] as follows:*

$$\lambda^{\star}_{\mathbf{\Phi},i} = \left[\eta - \frac{\sigma^2}{\lambda_{\mathbf{x},i}} \right]^{+}, \tag{2.39}$$

and $\eta > 0$ *is such that* $\sum_{i=1}^{\ell} \lambda^{\star}_{\mathbf{\Phi},i} \leq \ell$ *and* $[x]^{+} = max\{x, 0\}$*.*

Proof: See [93] . ■

We first present results applicable to a Gaussian source.

Theorem 18. *[93] Consider the linear measurement model in* (2.30)*, where* $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})$*, and* $\boldsymbol{\mu}_{\mathbf{x}}$ *is the mean and* $\boldsymbol{\Sigma}_{\mathbf{x}}$ *is the covariance matrix,* $s = rank(\boldsymbol{\Sigma}_{\mathbf{x}})$*. Then, we consider the measurement kernel* $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\star}$*, where* $\boldsymbol{\Phi}^{\star} \in \mathbb{R}^{\ell \times n}$ *solves* (2.37)*, it holds that*

1. *When* $\ell \geq s$ *the number of measurements is necessary to derive reconstruction MMSE to approach zero as* $\sigma^2 \to 0$*.*

2. *When* $\ell \geq s$ *the number of measurements is also sufficient to guarantee that MMSE converges to zero as* $\sigma^2 \to 0$*.*

Proof: See [93]. ■

The result presented in Theorem 18 is also to show that projection kernel design does not provide significant gains of reconstruction error in terms of number of measurements needed with respect to the random projection kernel as in Theorem 15.

We now present results applicable to a GMM source.

Theorem 19. *Consider the linear measurement model in* (2.30)*, where* $\mathbf{x} \sim \sum_k^K p_k \cdot \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(k)}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(k)})$*,* $s_{max} = max_k s_k$ *with* $s_k = rank(\boldsymbol{\Sigma}_{\mathbf{x}}^{(k)})$*, and* $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\star}$*. Then, by choosing the measurement kernel* $\boldsymbol{\Phi}^{\star} \in \mathbb{R}^{\ell \times n}$ *that solves* (2.37)*, it holds that*

$$\lim_{\sigma^2 \to 0} MMSE^{GM}(\sigma^2, \boldsymbol{\Phi}^{\star}) = 0 \Rightarrow \ell \geq s_{max}, \tag{2.40}$$

$$\ell > s_{max} \Rightarrow \lim_{\sigma^2 \to 0} MMSE^{GM}(\sigma^2, \boldsymbol{\Phi}^{\star}) = 0, \tag{2.41}$$

where $MMSE^{GM}(\sigma^2, \boldsymbol{\Phi}^{\star})$ *denotes the MMSE associated with the reconstruction of GMM signals using designed measurement kernel* $\boldsymbol{\Phi}^{\star}$*.*

Proof: See [93]. ■

The results embodied in Theorems 18 and 19 – appearing in [93] – are surprising because that they show that the number of measurements necessary to drive the MMSE to zero with a designed kernel are essentially the same as the number of measurements necessary to drive the MMSE to zero with a random kernel. We will see during the course of this thesis that this is not always the same in the presence of side information.

2.4.3 Reconstruction of GMM Signals with Side Information

We also overview the problem of reconstruction of high-dimensional signals from low-dimensional measurements in the presence of side information, by considering both the signal of interest and the side information signal are described by a joint GMM distribution [80, 81].

2.4.3.1 System Model

Let us consider the measurement model (see Fig. 2.5), where $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ is the signal of interest and $\mathbf{x}_2 \in \mathbb{R}^{n_2}$ is the side information. Let us also consider that the decoder has access to a set of linear measurements $\mathbf{y}_1 \in \mathbb{R}^{m_1}$ associated with the signal of interest, with $n_1 \geq m_1$:

$$\mathbf{y}_1 = \mathbf{\Phi}_1 \mathbf{x}_1 + \mathbf{w_1}; \tag{2.42}$$

and $\mathbf{y}_2 \in \mathbb{R}^{m_2}$ associated with the side information, with $n_2 \geq m_2$:

$$\mathbf{y}_2 = \mathbf{\Phi}_2 \mathbf{x}_2 + \mathbf{w_2}; \tag{2.43}$$

The matrices $\mathbf{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ and $\mathbf{\Phi}_2 \in \mathbb{R}^{m_2 \times n_2}$ represent the linear projection kernels for the target signal and the side information, respectively, and $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I_{n_1}} \cdot \sigma_1^2)$ and $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I_{n_2}} \cdot \sigma_2^2)$ are additive Gaussian noises that model possible distortion introduced by the feature extraction system.

Both signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are assumed to be described by a joint Gaussian mixture model. In particular, $\mathbf{x}_1$ is associated with underlying class labels $C_1 \in \{1, \cdots, K_1\}$, while $\mathbf{x}_2$ is associated with underlying class labels $C_2 \in \{1, \cdots, K_2\}$. The class labels obey the joint probability mass function (pmf) $P_{C_1,C_2}(i,k)$ and $\mathbf{x}_1$ and $\mathbf{x}_2$ conditioned on underlying class labels (C_1, C_2)=(i,k) obey the joint probability density function $p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k)$, so that $\mathbf{x}_1$ and $\mathbf{x}_2$ follow the

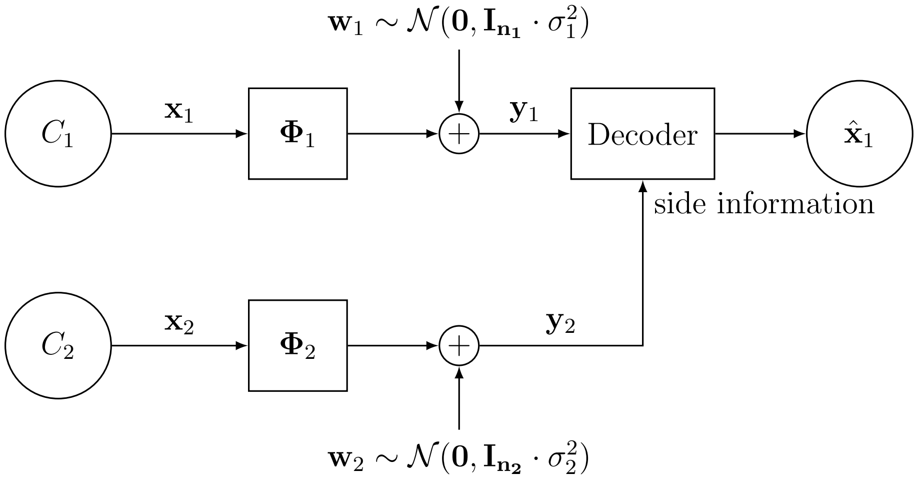


Figure 2.5: Compressive sensing model in the presence of side information.

GMM given by

$$\begin{aligned} p(\mathbf{x}_1, \mathbf{x}_2) &= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1,C_2}(i,k) p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k) \\ &= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1,C_2}(i,k) \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}), \end{aligned} \tag{2.44}$$

where

$$\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)} \\ \boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}, \quad \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)} & \mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \\ \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} & \mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}. \tag{2.45}$$

where $\boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)}$ and $\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)}$ are the mean and covariance matrix associated with the signal of interest $\mathbf{x}_1$, and likewise $\boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)}$ and $\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ are the mean and covariance matrix associated with the side information $\mathbf{x}_2$, respectively, and $\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}$ represents the cross-covariance between $\mathbf{x}_1$ and $\mathbf{x}_2$ given the classes $C_1 = i$ and $C_2 = k$.

We will also be assuming that the linear spaces associated to the images of the covariance matrices $\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)}$, $\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ and $\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}$ for different class labels $C_1 = i$ and $C_2 = k$ are independently drawn at random from a continuous pdf over the corresponding Grassmann manifold[4]. The covariance matrices are assumed to be possibly low-rank, such that $r_{\mathbf{x}} = \text{rank}(\mathbf{\Sigma}_{\mathbf{x}}) \leq n_1 + n_2$, $r_{\mathbf{x}_1} = \text{rank}(\mathbf{\Sigma}_{\mathbf{x}_1}) \leq n_1$ and

[4]Note that this assumption on the linear spaces occupied by signals in different classes reflects well the behavior of many real data ensembles for various applications such as face recognition, video motion segmentation, digits classification, etc. [91]. This assumption will enable us to simplify the statement of some of our results.

$r_{\mathbf{x}_2} = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_2}) \leq n_2$.

We are also interested in the reconstruction of the signal of interest $\mathbf{x}_1$ estimated via the noisy linear measurements $\mathbf{y}_1$ and $\mathbf{y}_2$ using the optimal (in MMSE sense) conditional estimator given by

$$\hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}[\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2] = \int_{-\infty}^{+\infty} \mathbf{x}_1 p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2) d\mathbf{x}_1, \tag{2.46}$$

where $p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2)$ is the *a posteriori* pdf of $\mathbf{x}_1$ given the measurements $\mathbf{y}_1$ and $\mathbf{y}_2$. Therefore, the reconstruction performance is measured in terms of the MMSE which is given by:

$$\text{MMSE}(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) = \mathbb{E}[||\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2)||^2], \tag{2.47}$$

Note that we express the MMSE explicitly as a function of the noise variances σ_1^2, σ_2^2 and projection kernels $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$.

We next report reconstruction results both in the scenario where both the signal of interest and the side information follow a Gaussian distribution ($K_1 = K_2 = 1$) and in the scenario where both the signal of interest and the side information follow a GMM distribution.

2.4.3.2 Gaussian Sources

We first consider conditions on the number of measurements that guarantee the reconstruction error to tend to zero in the low noise regime applicable to a Gaussian source.

Theorem 20. *[81] Consider the measurements model in (2.42) and (2.43), where $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a joint Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{x}}$, such that $r_{\mathbf{x}} = rank(\boldsymbol{\Sigma}_{\mathbf{x}})$, $r_{\mathbf{x}_1} = rank(\boldsymbol{\Sigma}_{\mathbf{x}_1})$ and $r_{\mathbf{x}_2} = rank(\boldsymbol{\Sigma}_{\mathbf{x}_2})$. Assume that measurement matrices $\boldsymbol{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ and $\boldsymbol{\Phi}_2 \in \mathbb{R}^{m_2 \times n_2}$ are randomly* *obtained drawn from a left-rotationally invariant distribution*[5]*. Then, we have*

$$\lim_{\sigma_1^2,\sigma_2^2 \to 0} MMSE(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 0 \Leftrightarrow m_1 \geq r_{\mathbf{x}_1} \ or \begin{cases} m_1 + m_2 \geq r_{\mathbf{x}} \\ m_1 \geq r_{\mathbf{x}} - r_{\mathbf{x}_2} \end{cases} . \quad (2.48)$$

Proof: See [81]. ■

Without side information, it is known that $m_1 \geq r_{\mathbf{x}_1}$ represents a necessary and sufficient condition on the number of measurements needed to guarantee that MMSE approaches 0 as $\sigma^2 \to 0$ as depicted in Theorem 15. On the other hand, it is possible to reconstruct the signal of interest $\mathbf{x}_1$ with lower number of measurements in the presence of side information as described in Theorem 20. In fact, whenever $r_{\mathbf{x}} < r_{\mathbf{x}_1} + r_{\mathbf{x}_2}$, it is possible to obtain accurate reconstruction of signal of interest in the low-noise regime even with less than $r_{\mathbf{x}_1}$ measurements, provided that $m_1 + m_2 \geq r_{\mathbf{x}}$. This happens when the number of measurements m_1 and m_2 have to be collectively greater than or equal to the dimension of the spaces spanned by signal $\mathbf{x}$. Moreover, the m_1 measurements need to be greater than or equal to the difference between the dimension of the spaces spanned by $\mathbf{x}$ and that spanned by the side information $\mathbf{x}_2$. It is also to note that the condition $m_1 \geq r_{\mathbf{x}_1}$ can also lead to reliable reconstruction as $\sigma^2 \to 0$, thus disregarding side information.

2.4.3.3 GMM Sources

We then consider conditions on the number of measurements that guarantee the reconstruction error to tend to zero in the low noise regime applicable to a GMM source.

Theorem 21. *[81] Consider the measurements model in (4.1) and (4.2), where* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *conditioned on the underlying class labels* $C_1 = i$, $C_2 = k$ *obey a joint Gaussian distribution with mean* $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}$ *and covariance* $\mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}$ *such that* $r_{\mathbf{x}}^{(i,k)} = rank(\mathbf{\Sigma}_{\mathbf{x}}^{(i,k)})$, $r_{\mathbf{x}_1}^{(i,k)} = rank(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)})$ *and* $r_{\mathbf{x}_2}^{(i,k)} = rank(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)})$ $\forall i, k$. *Assume*

[5] In this thesis, random $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are drawn from left-rotationally invariant distributions. A random matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ is said to be (left or right) rotation-invariant if the joint pdf of its entries $p(\mathbf{M})$ satisfies $p(\mathbf{\Theta M}) = p(\mathbf{M})$, or $p(\mathbf{M\Psi}) = p(\mathbf{M})$, respectively, for any orthogonal matrix $\mathbf{\Theta}$ or $\mathbf{\Psi}$. A special case of (left and right) rotation-invariant random matrices is represented by matrices with independent identically distributed (i.i.d.), zero-mean Gaussian entries with fixed variance.

that measurement matrices $\mathbf{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ and $\mathbf{\Phi}_2 \in \mathbb{R}^{m_2 \times n_2}$ are randomly constituted drawn from a left-rotationally invariant distribution. Then, the number of projections m_1 and m_2 that are sufficient to derive the MMSE to zero are given by

$$m_1 > r_{\mathbf{x}_1}^{(ik)} \quad \text{or} \quad \begin{cases} m_1 > r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_2}^{(ik)} \\ m_1 + m_2 > r_{\mathbf{x}}^{(ik)} \end{cases} \Rightarrow \lim_{\sigma_1^2, \sigma_2^2 \to 0} MMSE(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 0. \tag{2.49}$$

and necessary conditions are given by

$$\lim_{\sigma_1^2, \sigma_2^2 \to 0} MMSE(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) \Rightarrow m_1 \geq r_{\mathbf{x}_1}^{(ik)} \quad \text{or} \quad \begin{cases} m_1 \geq r_{\mathbf{x}}^{(ik)} - r_{\mathbf{x}_2}^{(ik)} \\ m_1 + m_2 \geq r_{\mathbf{x}}^{(ik)} \end{cases}, \tag{2.50}$$

for all (i,k) $\in \mathcal{S}$.

Proof: See [81]. ■

The necessary condition presented in Theorem 21 shows that – akin to the Gaussian case – the numbers of measurements m_1 and m_2 need to be greater than or equal to the largest among the dimensions of the spaces spanned by $\mathbf{x}$ in the Gaussian components corresponding to class labels $(C_1, C_2) = (i, k)$, for $i = \{1, \cdots, K_1\}$ and $k = \{1, \cdots, K_2\}$. Moreover, the m_1 measurements would need to be enough to span a space with dimension equal to the difference between the dimension of the spaces spanned by $\mathbf{x}$ and that spanned by the side information $\mathbf{x}_2$. We can also observe that the condition $m_1 \geq r_{\mathbf{x}_1}$ can guarantee the perfect reconstruction of the signal of interest as $\sigma^2 \to 0$, thus disregarding side information. Furthermore, it would be of interest to note that the sufficient conditions for reliable reconstruction of GMM sources are one feature away from the corresponding necessary conditions, akin to previous results in Theorem 16 for the case of no side information.

Theorems 20 and 21 determine sharp necessary and sufficient conditions on the number of measurements needed for the MMSE to approach zero as in the low-noise regime, where such projection kernels $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are randomly constituted drawn from a left-rotationally invariant distribution. This thesis will be asking the question about whether it is possible to improve further the performance of compressive sensing systems with side information by appropriate design of the projection kernels $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$.

2.5 Summary

This chapter has reviewed the major advances associated with compressive sensing. We first introduced the classical Nyquist–Shannon sampling theorem as well as recent advances in sampling theory that have motivated compressive sensing. We also presented compressive sensing framework, including the signal sensing and acquiring process. In particular, we focused on how to construct the measurement matrix and how to accurately reconstruct a signal in the compressive sensing system. Furthermore, we discussed the most commonly used reconstruction algorithms with their corresponding recovery guarantees such as the convex ℓ_1-norm optimization and other iterative greedy algorithms.

We also introduced, in this chapter, the extensions of compressive sensing applications, including the use of side information. In particular, we presented advanced works in compressive sensing that leverage the presence of side information in various application scenarios such as medical imaging and remote sensing. We also studied further about the sensing model and reconstruction algorithms, as well as their corresponding recovery guarantees in the presence of side information.

Finally, we focused on the compressive sensing systems that leverage other structured signal models including union of subspaces, wavelet trees, and manifolds, that aim to capture additional signal structure beyond signal sparsity. We then studied one structured model – the GMM – which can be used in conjunction with compressive sensing systems. In addition, we investigated the problem of reconstruction of GMM signals in the CS with/without side information. We also provided sharp sufficient and necessary conditions on the number of linear projections for reliable reconstruction in the low-noise regime for both side information scenarios.

The compressive sensing results applicable to GMM signal models motivate the work reported in this thesis.

Chapter 3

Measurement Design with Side Information – Capturing the Signal of Interest

In this chapter, we introduce our proposed CS model that leverages the presence of side information. In this work, we focus on the study of the reconstruction performance associated with measurements design, by assuming that both the signal of interest and the side information are drawn from a joint Gaussian mixture model (GMM) distribution. We then develop two different scenarios of measurements designs associated to the signal of interest, by considering that i) the side information is available at the decoder only; and ii) the case of side information at both the decoder and encoder.

In particular, we consider projection designs aimed at minimizing the reconstruction minimum mean-squared error (MMSE) and that leverage the presence of side information. We determine sharp sufficient and necessary conditions on the number of measurements needed to guarantee that the MMSE approaches zero in the low-noise regime. Finally, numerical results with both synthetic data and real imaging are presented to confirm our proposed theorems and bounds aligned well with practice. At the end of this chapter, we summarize the main contributions in this work.

3.1 Problem Statement

3.1.1 Model

We study the problem associated with the reconstruction of high-dimensional signals (signal of interest) from low-dimensional measurements in the presence of side information, as depicted in Fig. 3.1. In particular, we consider the case when side information is presented at the decoder (and possibly at the encoder) which is uncompressed without additive noise.

We are interested in the reconstruction of the signal of interest $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ from noisy, linear, compressive measurements $\mathbf{y}_1 \in \mathbb{R}^{m_1}$, with $m_1 \leq n_1$, given by:

$$\mathbf{y}_1 = \mathbf{\Phi}_1 \mathbf{x}_1 + \mathbf{w}_1, \tag{3.1}$$

where the matrix $\mathbf{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ represents the linear projection kernel and $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma_1^2)$ is the additive white Gaussian noise that models possible distortion introduced in the sensing process and ambient noise, and from side information $\mathbf{x}_2 \in \mathbb{R}^{n_2}$.

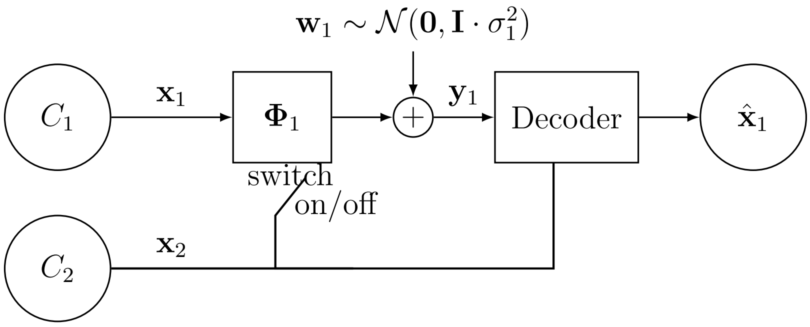


Figure 3.1: Compressive sensing model in the presence of side information.

We are also interested in reconstruction of the signal of interest $\mathbf{x}_1$ from the noisy, linear measurements $\mathbf{y}_1$ and side information $\mathbf{x}_2$ using the optimal (in MMSE sense) conditional estimator given by

$$\hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{x}_2) = \mathbb{E}[\mathbf{x}_1 | \mathbf{y}_1, \mathbf{x}_2] = \int_{-\infty}^{+\infty} \mathbf{x}_1 p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{x}_2) d\mathbf{x}_1, \tag{3.2}$$

where $p(\mathbf{x}_1|\mathbf{y}_1, \mathbf{x}_2)$ is the *a posteriori* pdf of $\mathbf{x}_1$ given the compressive measurements $\mathbf{y}_1$ and the side information $\mathbf{x}_2$.

Therefore, the reconstruction performance is measured in terms of the MMSE which is given by

$$\mathrm{MMSE}(\sigma_1^2, \mathbf{\Phi}_1) = \mathbb{E}[||\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{x}_2)||^2], \tag{3.3}$$

which is a function of the joint distribution $p(\mathbf{x}_1, \mathbf{x}_2)$, the projection kernel $\mathbf{\Phi}_1$ and the noise variance σ_1^2.

Finally, we consider that the decoder and the encoder have access to the joint pdf of the signal of interest and the side information $p(\mathbf{x}_1, \mathbf{x}_2)$. We also consider that the signal of interest $\mathbf{x}_1$ and the side information signal $\mathbf{x}_2$ – which are correlated – are described by a joint Gaussian mixture model. In particular, $\mathbf{x}_1$ is associated with underlying class labels $C_1 \in \{1, \ldots, K_1\}$ and $\mathbf{x}_2$ is associated with underlying class labels $C_2 \in \{1, \ldots, K_2\}$, respectively. The class labels obey the joint pmf $P_{C_1,C_2}(i,k)$ and the signals $\mathbf{x}_1$ and $\mathbf{x}_2$ conditioned on the underlying class labels $C_1 = i$ and $C_2 = k$, obey the joint Gaussian pdf:

$$p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}), \tag{3.4}$$

and so $\mathbf{x}_1$ and $\mathbf{x}_2$ follow the GMM given by

$$\begin{aligned} p(\mathbf{x}_1, \mathbf{x}_2) &= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1,C_2}(i,k) p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k) \\ &= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1,C_2}(i,k) \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}), \end{aligned} \tag{3.5}$$

where

$$\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)} \\ \boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}, \quad \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)} & \mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \\ \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} & \mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}. \tag{3.6}$$

Here, $\boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)}$ and $\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)}$ are the mean and covariance matrix of $\mathbf{x}_1$, conditioned on class labels $C_1 = i, C_2 = k$, and likewise $\boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)}$ and $\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ are the mean and covariance matrix of $\mathbf{x}_2$, conditioned on class labels $C_1 = i, C_2 = k$, respectively. The cross-covariance between $\mathbf{x}_1$ and $\mathbf{x}_2$ given the class labels $C_1 = i$ and $C_2 = k$ is given by $\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}$.

We will also be assuming that the linear spaces associated to the images of the covariance matrices $\mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}$, $\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)}$ and $\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ associated to different class labels $C_1 = i$ and $C_2 = k$ are independently drawn at random from a continuous pdf over the corresponding Grassmann manifold.

3.2 Design Method

We now introduce our measurement design problem. In particular, we consider measurement design in two scenarios: i) side information is available only *at the decoder*, i.e., the optimization of the projection kernel $\mathbf{\Phi}_1$ assumes that the realization of the side information is available at the decoder only; ii) side information is available *at the encoder and the decoder*, i.e., the optimization of the projection kernel $\mathbf{\Phi}_1$ assumes that the realization of the side information is available at both the decode and encoder. We also consider, for comparison purpose, the scenario where the linear measurement kernel $\mathbf{\Phi}_1$ is randomly constituted[1].

We now discuss two different scenarios for kernel design, based on the fact that the side information $\mathbf{x}_2$ is offered only to the decoder (i.e., by considering the switch associated to the encoder is "off" in Fig. 3.1), or both at decoder and encoder (i.e., by considering the switches associated to the encoder is "on" in Fig. 3.1):

1. *Side information available at the decoder only*: when side information is available at the decoder only, we aim to determine the linear projection kernel that minimizes the MMSE in (3.3) subject to appropriate constraints. In this case, the designed kernel $\mathbf{\Phi}_1^\star$ corresponds to the solution of the optimization problem

$$\begin{aligned} &\underset{\mathbf{\Phi}_1}{\text{minimize}}\ \mathrm{MMSE}(\sigma_1^2, \mathbf{\Phi}_1) \\ &\text{subject to}\ \mathrm{tr}(\mathbf{\Phi}_1 \mathbf{\Phi}_1^{\mathrm{T}}) \le m_1, \end{aligned} \tag{3.7}$$

where the trace constraint in (3.7) limits the average energy associated to the projection kernel. We denote the MMSE associated with the optimal

[1]Note that the random measurement matrix $\mathbf{\Phi}_1$ is drawn from left-rotationally invariant distributions as in Chapter 2.

kernel design for this case as follows:

$$\mathrm{MMSE}_{\mathsf{d}}^{\mathsf{opt}}(\sigma_1^2) = \mathrm{MMSE}(\sigma_1^2, \mathbf{\Phi}_1^\star). \tag{3.8}$$

2. *Side information available both at encoder and decoder*: on the other hand, when side information is available both at encoder and decoder, we aim to determine the linear measurement kernel that minimize the conditional MMSE

$$\mathrm{MMSE}(\sigma_1^2, \mathbf{\Phi}_1, \mathbf{x}_2) = \mathbb{E}[\|\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{x}_2)\|^2 | \mathbf{x}_2], \tag{3.9}$$

In this case, the designed kernel $\mathbf{\Phi}_1^\star(\mathbf{x}_2)$ – which is a function of the realization of side information $\mathbf{x}_2$ – correspond to the solution of the optimization problem

$$\begin{aligned} &\underset{\mathbf{\Phi}_1}{\mathrm{minimize}}\ \mathrm{MMSE}(\sigma_1^2, \mathbf{\Phi}_1, \mathbf{x}_2) \\ &\text{subject to } \mathrm{tr}(\mathbf{\Phi}_1 \mathbf{\Phi}_1^{\mathrm{T}}) \le m_1. \end{aligned} \tag{3.10}$$

We denote the MMSE associated with the optimal kernel design for this case as

$$\mathrm{MMSE}_{\mathsf{ed}}^{\mathsf{opt}}(\sigma_1^2) = \mathbb{E}[\mathrm{MMSE}(\sigma_1^2, \mathbf{\Phi}_1^\star(\mathbf{x}_2), \mathbf{x}_2)], \tag{3.11}$$

where the expectation is taken with respect to $\mathbf{x}_2$.

Note that, in all scenarios, the decoder has perfect knowledge of the joint pdf $p(\mathbf{x}_1, \mathbf{x}_2)$, the projection kernel $\mathbf{\Phi}_1$ adopted to compress $\mathbf{x}_1$, and the noise variance $\boldsymbol{\sigma}_1^2$.

Our objective is to determine the necessary and sufficient conditions on the number of measurements needed for the reconstruction error to approach zero in the low noise regime, i.e.

$$\lim_{\sigma_1^2 \to 0} \mathrm{MMSE}(\sigma_1^2, \mathbf{\Phi}_1) = 0. \tag{3.12}$$

Such conditions will be shown both for the scenario where the measurement matrix $\mathbf{\Phi}_1$ is random, and for the scenario where measurement matrix associated to the signal of interest is designed when $\mathbf{\Phi}_1 = \mathbf{\Phi}_1^\star$, in case where side information is available at decoder only, and at encoder and decoder.

We can note that it is possible to express the conditional mean estimator (3.2) in closed-form but not the MMSE (3.3), which lead to the difficulty in determining the characterization of such bounds on the number of measurements needed for reliable reconstruction (i.e., such that (3.12) holds) when the solutions of optimization problem in (3.7) and (3.10) cannot be analytically characterized.

We also note that the difference between this work and the previous work in [81] related to the fact that [81] analyses the performance of compressive sensing systems with side information with random measurement strategies, whereas this work analyses the performance of these systems with designed measurement strategies.

3.3 Bounds on the Number of Measurements for Reliable Reconstruction

In this section, we study the impact of measurement design by analyzing low-noise MMSE *phase transition*. Namely, a phase transition in the reconstruction MMSE is observed when the number of measurements extracted from the signal of interest m_1 is such that the corresponding MMSE approaches zero when noise $\sigma_1^2 \to 0$. We now focus on determining sufficient and necessary conditions on the minimum number of measurements needed for reliable reconstruction (i.e. such that (3.12) holds).

3.3.1 Gaussian Source

We start by considering our proposed measurement model in (3.1) with side information $\mathbf{x}_2$, by assuming that $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a jointly Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance matrix $\mathbf{\Sigma}_{\mathbf{x}}$, i.e., when $K_1 = K_2 = 1$ in (3.6). The covariance matrices are assumed to be possibly low-rank, such that $r_{\mathbf{x}} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{x}}) \leq n_1 + n_2$, $r_{\mathbf{x}_1} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{x}_1}) \leq n_1$ and $r_{\mathbf{x}_2} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{x}_2}) \leq n_2$.

Theorem 22. *Consider the case when side information $\mathbf{x}_2$ is available at the decoder only, with $\mathbf{\Phi}_1 = \mathbf{\Phi}_1^\star$, where $\mathbf{\Phi}_1^\star$ is obtained via through the optimization* *problem* (3.7)*. It holds*

$$\lim_{\sigma_1^2 \to 0} MMSE_{\mathsf{d}}^{\mathsf{opt}}(\sigma_1^2) = 0 \Leftrightarrow m_1 \geq \ r_{\mathbf{x}} - r_{\mathbf{x_2}}, \tag{3.13}$$

Proof: See Appendix A.1. ∎

Theorem 23. *Consider the case when side information* $\mathbf{x}_2$ *is available at both encoder and decoder, with* $\mathbf{\Phi}_1 = \mathbf{\Phi}_1^\star$*, where* $\mathbf{\Phi}_1^\star$ *is obtained via through the optimization problem* (3.10)*. Then,*

$$\lim_{\sigma_1^2 \to 0} MMSE_{\mathsf{ed}}^{\mathsf{opt}}(\sigma_1^2) = 0 \Leftrightarrow m_1 \geq \ r_{\mathbf{x}} - r_{\mathbf{x_2}}, \tag{3.14}$$

Proof: See Appendix A.1. ∎

The proposed conditions in Theorems 22 and 23 show that the same necessary and sufficient conditions for the MMSE phase transition hold for both designed cases when side information is available at the decoder only, or at both decoder and encoder. We can observe that the projection design for the signal of interest does not provide significant gains of reconstruction error in terms of number of measurements needed with respect to such conditions in Theorem 20 where projection kernels are random.

In addition to this, the necessary and sufficient conditions for the MMSE approaches to zero in the low-noise regime suggest that the number of measurements m_1 must be greater than or equal to $r_{\mathbf{x}} - r_{\mathbf{x}_2}$. This means the m_1 measurements need to be greater than or equal to the difference between the dimension of the spaces spanned by $\mathbf{x}$ and that spanned by the side information $\mathbf{x}_2$. We also recall the case with no side information, where we need $m_1 \geq r_{\mathbf{x}_1}$ to achieve the reliable reconstruction as depicted in Theorem 15. This means the availability of side information in CS model can lead to significant gains of reconstruction performance in terms of the number of measurements for the MMSE to approach zero in the low-noise regime, as already observed in [81].

3.3.2 GMM Sources

We now study the case where $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint GMM in (3.5), characterized by the underlying class labels $C_1 = i, C_2 = k$, with mean $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$, $\forall i, k$. The covariance matrices are assumed to be possibly low-rank, where $r_{\mathbf{x}}^{(i,k)} = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}) \leq n_1 + n_2$, $r_{\mathbf{x}_1}^{(i,k)} = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}) \leq n_1$ and $r_{\mathbf{x}_2}^{(i,k)} = \text{rank}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}) \leq n_2$.

Theorem 24. *Consider the case when side information $\mathbf{x}_2$ is available at the decoder only, with $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_1^\star$, where $\boldsymbol{\Phi}_1^\star$ is obtained via through the optimization problem* (3.7)*. Then,*

$$\lim_{\sigma_1^2 \to 0} MMSE_{\mathsf{d}}^{\mathsf{opt}}(\sigma_1^2) = 0 \Rightarrow m_1 \geq \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)}, \tag{3.15}$$

$$m_1 > \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)} \Rightarrow \lim_{\sigma_1^2 \to 0} MMSE_{\mathsf{d}}^{\mathsf{opt}}(\sigma_1^2) = 0. \tag{3.16}$$

Proof: See Appendix A.2. ■

Theorem 25. *Consider now the case when side information $\mathbf{x}_2$ is available at both encoder and decoder, with $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_1^\star(\mathbf{x}_2)$, where, for each value of $\mathbf{x}_2$, $\boldsymbol{\Phi}_1^\star(\mathbf{x}_2)$ is the solution of the optimization problem* (3.10)*. Then,*

$$\lim_{\sigma_1^2 \to 0} MMSE_{\mathsf{ed}}^{\mathsf{opt}}(\sigma_1^2) = 0 \Rightarrow m_1 \geq \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)}, \tag{3.17}$$

$$m_1 > \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)} \Rightarrow \lim_{\sigma_1^2 \to 0} MMSE_{\mathsf{ed}}^{\mathsf{opt}}(\sigma_1^2) = 0. \tag{3.18}$$

Proof: See Appendix A.2. ■

The results in Theorems 24 and 25 show that the same necessary and sufficient conditions for the MMSE phase transition hold for both designed cases. The necessary conditions presented in Theorems 24 and 25 suggest that the number of measurements m_1 needs to be greater than or equal to the largest among the difference between the dimension of the spaces spanned by $\mathbf{x}$ and that spanned by side information $\mathbf{x}_2$ in the Gaussian components corresponding to class labels $(C_1, C_2) = (i, k)$, for $i = \{1, \cdots, K_1\}$ and $k = \{1, \cdots, K_2\}$. Furthermore, we can also observe that the sufficient conditions for reliable reconstruction of GMM sources are one feature away from the corresponding necessary conditions.

Moreover, on comparing such conditions with those in Theorem 21 obtained for the case of random measurements, the projection design associated to the signal of interest does not reduce the minimum number of measurements with respect to the random one for reliable reconstruction. It is also important to note that even though all cases require the same number of measurements for phase transition, measurements design can still lead to gains of finite SNR.

3.4 Numerical Results

In this section, we provide numerical results to showcase that our analysis is aligned with practice. We first report results for synthetic data, considering both cases of Gaussian sources and GMM sources, that highlighting the value of using an optimal measurement matrix $\mathbf{\Phi}_1{}^\star$ with respect to a random one. We then provide reconstruction results with real imaging data to showcase the impact of projection design.

3.4.1 Synthetic Data: Gaussian Sources

We start by considering the case when $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint Gaussian distribution with $K_1 = K_2 = 1$. We assume that signal dimensions are $n_1 = 10$ and $n_2 = 6$. Both signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are zero-mean, i.e., $\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{0}$, and the corresponding covariance matrices are independently drawn from a continuous pdf over the corresponding Grassmann manifold such that $r_{\mathbf{x}} = 5$, $r_{\mathbf{x}_1} = 3$ and $r_{\mathbf{x}_2} = 3$. In this experiment, we consider the projection kernel $\mathbf{\Phi}_1$ is randomly obtained with i.i.d., zero-mean, Gaussian entries with fixed variance, whereas designed kernels are obtained by approximating numerically[2] the solution of the problems in (3.7) and (3.10).

It is possible to predict that the MMSE approaches zero in the low-noise regime with number of measurements $m_1 \geq r_{\mathbf{x}_1} = 3$ for the case of no side information via the analysis reported in Theorem 15. On the other hand, Fig. 3.3 shows the values of reconstruction MMSE for the case of side information available at the decoder (with random and designed kernels) and at both the encoder and decoder. The numerical results are perfectly aligned with the predictions

[2]We consider the gradient descent optimization algorithm for finding the minimum error iteratively.

contained in Theorems 22 and 23, as we observe that the MMSE approaches zero in the low-noise regime with $m_1 \geq r_{\mathbf{x}} - r_{\mathbf{x_2}} = 2$, for both cases of random and designed projection kernels. We can also observe that both designs in (3.7) and (3.10) do not provide advantage in terms of MMSE phase transition, but surprisingly they have a significant impact on the reconstruction performance (almost 20 dB) for the finite noise levels. In particular, the design in (3.10) does not provide significant gains of reconstruction MMSE with respect to the design in (3.7) (see cross and circle lines in Fig. 3.3).

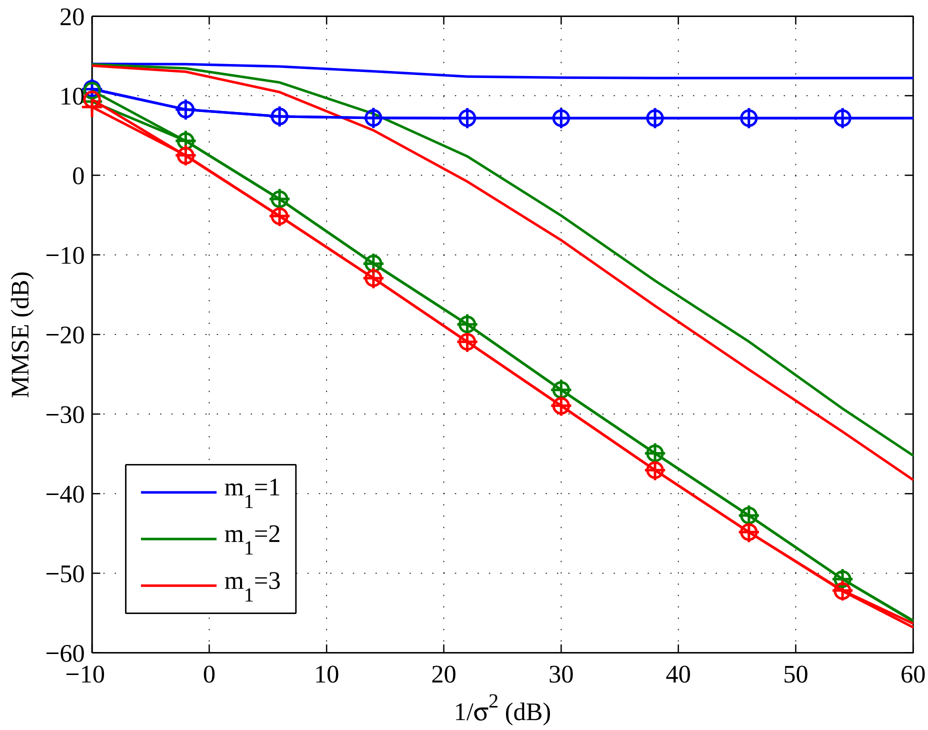


Figure 3.2: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ with side information. Side information at the decoder only with random projection kernel (solid lines) and with designed projection kernel (cross lines). Side information at both the encoder and the decoder with designed projection kernel (circles).

We then consider a joint Gaussian distribution for signals $\mathbf{x}_1$ and $\mathbf{x}_2$ with $K_1 = K_2 = 1$, by noting that they are high-dimensional signals, i.e. $n_1 = 100$ and $n_2 = 20$. All means associated with the various Gaussian distributions are zero, i.e. $\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{0}$, and the covariance matrices are randomly generated such that $r_{\mathbf{x}} = 5$, $r_{\mathbf{x}_1} = 3$ and $r_{\mathbf{x}_2} = 3$. As used in the previous experiment, we consider the projection kernel $\mathbf{\Phi}_1$ is randomly obtained with i.i.d., zero-mean, Gaussian entries with fixed variance, whereas designed kernels are obtained by approximating numerically the solution of the problems in (3.7) and (3.10).

Fig. 3.3 illustrates the values of reconstruction MMSE for the case of side information available at the decoder (with random and designed kernels) and at both the encoder and decoder. We can observe that the MMSE tends to zero when $m_1 \geq 2$ in the low-noise regime, for both cases of random (solid lines) and designed (cross and circle lines) projection kernels, as predicted by Theorems 22 and 23. In this experiment, we can also observe that both designs in (3.7) and (3.10) have significant impact on the reconstruction performance for the finite noise levels even though they do not provide MMSE phase transition. Finally, we can also note that our theoretical results are well aligned with numerical ones for high-dimensional signals $\mathbf{x}_1$ and $\mathbf{x}_2$.

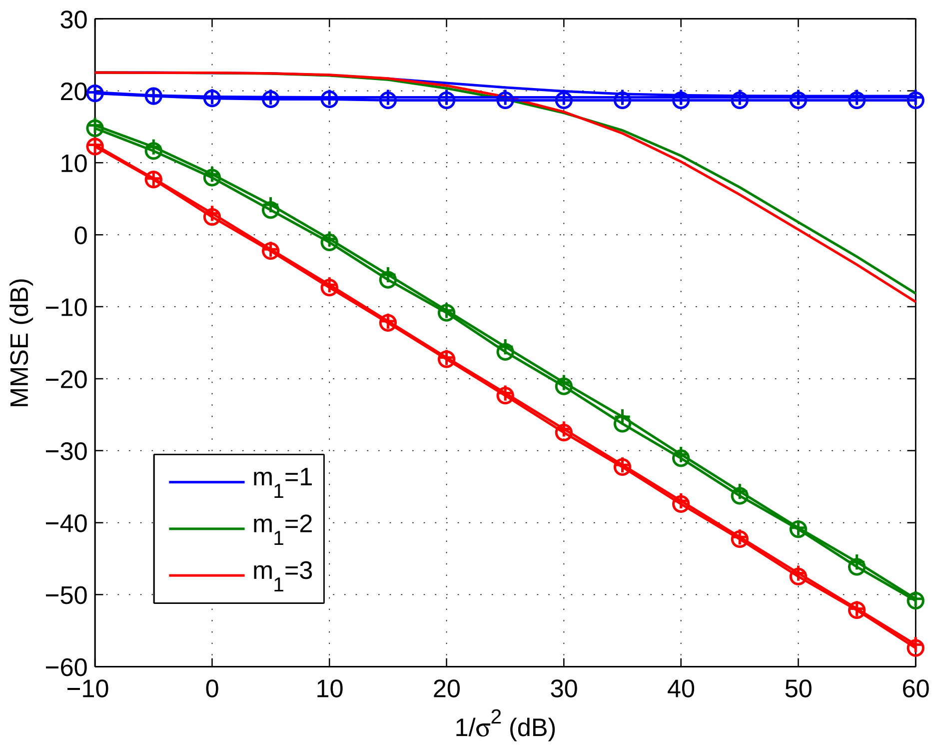


Figure 3.3: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ with side information for high-dimensional signals $\mathbf{x}_1$ and $\mathbf{x}_2$. Side information at the decoder only with random projection kernel (solid lines) and with designed projection kernel (cross lines). Side information at both the encoder and the decoder with designed projection kernel (circles).

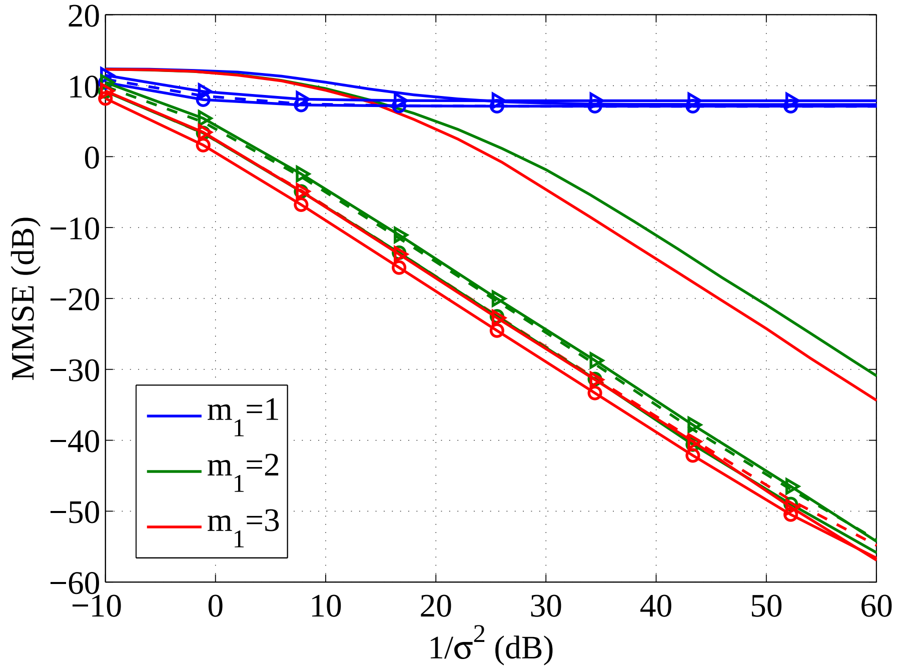


Figure 3.4: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ with side information. Side information at the decoder only with random projection kernel (solid lines) and with designed projection kernel (dashed lines). Side information at both the encoder and the decoder with designed projection kernel (circles) and suboptimal design (classification of $\mathbf{x}_2$) (triangles).

3.4.2 Synthetic Data: GMM Sources

We now consider a joint GMM distribution for signals $\mathbf{x}_1$ and $\mathbf{x}_2$ with dimensions $n_1 = 10$ and $n_2 = 6$, and $K_1 = K_2 = 2$. Both signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are zero-mean, i.e., $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \mathbf{0}$, and the covariance matrices associated to different classes are drawn uniformly at random from the corresponding Grassmann manifold, such that $r_{\mathbf{x}}^{(i,k)} = 5$, $r_{\mathbf{x}_1}^{(i,k)} = 3$ and $r_{\mathbf{x}_2}^{(i,k)} = 3$.

Fig. 3.4 illustrates the reconstruction MMSE vs. $1/\sigma^2$ for synthetic data. By using the results in Theorem 16, we can predict that phase transition of the MMSE is obtained only when number of measurements $m_1 > \max_{i,k} r_{\mathbf{x}_1}^{(i,k)}$ for the case when no side information is available in [93]. On the other hand, the impact of side information is showcased in Fig. 3.4, where we report the MMSE values for the case of side information available at the decoder (with random and designed kernels) and at both encoder and decoder. More precisely, we assume that the random projection kernel $\mathbf{\Phi}_1$ is generated with i.i.d., zero-mean, Gaussian entries with fixed variance, whereas designed kernels are obtained by approximating numerically the solution of the problems in (3.7) and (3.10).

We also consider a further projection design scheme for the case of side information at the encoder and decoder. First, class labels $\hat{C}_1$ and $\hat{C}_2$ are estimated via MAP classifier from $\mathbf{y}_1$ and the side information $\mathbf{x}_2$. Then, $\mathbf{\Phi}_1$ is obtained in (2.38) as the optimal kernel design for Gaussian inputs with distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_1}^{(\hat{C}_1,\hat{C}_2)}, \mathbf{\Sigma}_{\mathbf{x}_1}^{(\hat{C}_1,\hat{C}_2)})$, as described in Theorem 17, thus allowing to reduce such computationally expensive numerical solution of the optimization problem in (3.7) and (3.10).

In Fig. 3.4, we can observe that the numerical results are well aligned with our Theorems 24 and 25, when the MMSE tends to zero in the low-noise regime when number of measurements $m_1 > \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} = 2$, for the case of random projection kernels, and as well as the case of designed kernels. Moreover, careful design of the projection kernel associated to the signal of interest in (3.7) and the availability of side information at the encoder side in (3.10) do not provide a further advantage in terms of measurements needed for the MMSE phase transition with respect to the random one, but they provide significant gains (almost 30 dB) in terms of robustness against noise. We can also observe that kernel design schemes that leverage the presence of side information at the encoder and the decoder do not provide significant advantages in comparison to kernel designs with side information at the decoder only. In particular, we also note that the suboptimal scheme for the case of side information at both encoder and decoder, though computationally simpler than the design obtained via the solution of the problem in (3.10), provides very similar MMSE values (see circle lines and triangle lines in Fig. 3.4).

We then consider a joint GMM distribution for signals $\mathbf{x}_1$ and $\mathbf{x}_2$ with higher dimensions $n_1 = 100$ and $n_2 = 20$, and $K_1 = K_2 = 2$. Both signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are zero-mean, i.e., $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \mathbf{0}$, and the covariance matrices associated to different classes are drawn uniformly at random from the corresponding Grassmann manifold, such that $r_{\mathbf{x}}^{(i,k)} = 5$, $r_{\mathbf{x}_1}^{(i,k)} = 3$ and $r_{\mathbf{x}_2}^{(i,k)} = 3$.

In this experiment, we consider the projection kernel $\mathbf{\Phi}_1$ is randomly obtained with i.i.d., zero-mean, Gaussian entries with fixed variance, whereas designed

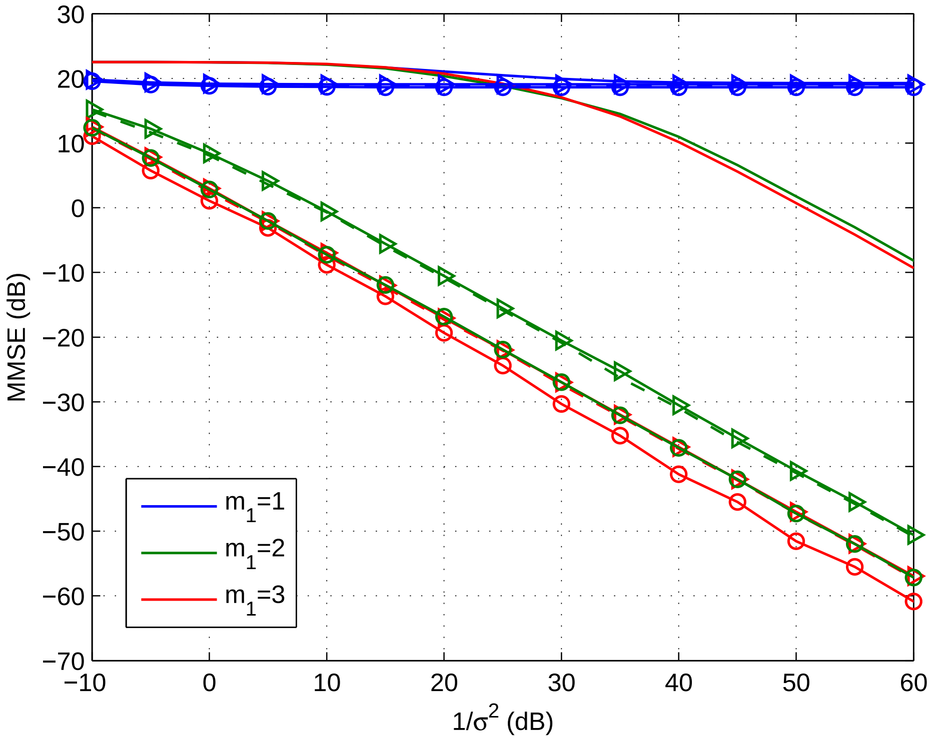


Figure 3.5: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ with side information for high-dimensional signals $\mathbf{x}_1$ and $\mathbf{x}_2$. Side information at the decoder only with random projection kernel (solid lines) and with designed projection kernel (dashed lines). Side information at both the encoder and the decoder with designed projection kernel (circles) and suboptimal design (classification of $\mathbf{x}_2$) (triangles).

kernels are obtained by approximating numerically the solution of the problems in (3.7) and (3.10). Moreover, we also consider a further projection design scheme for the case of side information at the encoder and decoder, by assuming that $\mathbf{\Phi}_1$ is obtained in (2.38) as the optimal kernel design for Gaussian inputs with distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_1}^{(\hat{C}_1,\hat{C}_2)}, \mathbf{\Sigma}_{\mathbf{x}_1}^{(\hat{C}_1,\hat{C}_2)})$, as described in Theorem 17, where class labels $\hat{C}_1$ and $\hat{C}_2$ are estimated via MAP classifier from $\mathbf{y}_1$ and the side information $\mathbf{x}_2$.

Fig. 3.5 reports the MMSE vs. $1/\sigma^2$. We can note that the MMSE tends to zero in the low-noise regime when number of measurements $m_1 > \max_{i,k} r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} = 2$, for the case of random projection kernels (solid lines), and as well as the case of designed kernels (dashed, circle and triangle lines). In particular, we also observe that the suboptimal scheme for the case of side information at both encoder and decoder, achieves similar PSNR values as when the projection kernel is designed via more computationally expensive numerical solution of the optimization problem in (3.10) (see circle and triangle lines). Moreover, the kernel design associated to the signal of interest in (3.7) and the availability of side information at the encoder side in (3.10) do not reduce the number of measurements need for the MMSE phase transition but they provide significant gains in terms of robustness against noise with respect to the random ones. Finally, we can also note that our proposed theorems are well aligned with numerical ones for high-dimensional signals $\mathbf{x}_1$ and $\mathbf{x}_2$.

3.4.3 Real Data

We now consider a reconstruction example with real imaging data to see whether the measurement design problem posed in (3.7) and (3.10) lead to better reconstruction performance. We use a high-resolution image "Lena" with resolution 512×512 as the signal of interest and a low-resolution version of the same subject (128×128 pixels) as side information (See Fig. 3.6). In this work, both images are assumed to be divided into the same number of patches, which correspond to the same spatial portion of the subject. Therefore, we consider both signals are partitioned into non-overlapping patches, so that vectors $\mathbf{x}_1$ represent 8×8 non-overlapping patches extracted from the signal of interest, and vectors $\mathbf{x}_2$ represent 2×2 non-overlapping patches extracted from the side information, that are partitioned into overall 4096 patches for both images.

The vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are assumed to be described by a joint GMM prior with $K_1 = K_2 = 20$ classes. The parameters of the joint GMM, i.e., prior class probabilities, class-conditioned means and class-conditioned covariance matrices, are trained using the expectation-maximization (EM) algorithm [96] over a set of patches extracted from images in the Caltech 101 dataset [97]. We note that "Lena" is not in the training ensemble. In this experiment, we consider the designed projection kernels are obtained by approximating numerically the solution of the problems in (3.7) and (3.10). We also consider a projection kernel $\mathbf{\Phi}_1$ that is randomly obtained with i.i.d., zero-mean, Gaussian entries with fixed variance,

Figure 3.6: Left: input signal, high-resolution image (512×512). Right: side information, low-resolution image (128×128)

for comparison purpose.

In Fig. 3.7, we report some reconstruction examples obtained with $m_1 = 15$ linear measurements from each non-overlapping patch, and with noise level $\sigma^2 = -40$ dB. From the top to the bottom, the reconstruction images correspond the following different scenarios: (a) random projection kernel with i.i.d., zero-mean, Gaussian entries with fixed variance with side information at the decoder only; (b) designed projection kernel with side information at the decoder only; (c) designed kernel with side information at both the encoder and the decoder. In the last case, we have considered the design based on the classification of $\mathbf{y}_1$ and the side information $\mathbf{x}_2$ on the kernel construction for single Gaussian sources described in Theorem 17. In particular, the design $\mathbf{\Phi}_1$ is obtained in (2.38) as the optimal kernel design for Gaussian inputs with distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}_1}^{(\hat{C}_1,\hat{C}_2)}, \mathbf{\Sigma}_{\mathbf{x}_1}^{(\hat{C}_1,\hat{C}_2)})$, where class labels $\hat{C}_1$ and $\hat{C}_2$ are estimated via a MAP classifier from the vectors $\mathbf{y}_1$ and $\mathbf{x}_2$, thus allowing to reduce such computationally expensive numerical solution of the optimization problem in (3.7) and (3.10).

The reconstruction results illustrate that both designs (as shown in Figs. 3.7(b) and (c)) can significantly improve reconstruction quality, which lead to peak signal-to-noise ratio (PSNR) gains of approximately 6 dB with respect to the random projection kernel in Fig. 3.7(a). In this experiment, we observe that providing the encoder with side information does not guarantee significant advantages. We can also note that suboptimal approach based on the classification of $\mathbf{y}_1$ and $\mathbf{x}_2$ yields PSNR values slightly lower than those obtained with side information at the decoder only.

We now consider different number of measurements to assess the impact of projection kernel design. In this experiment, we also use the image "Lena" 512×512 as the signal of interest and a low-resolution version of the same subject (128×128 pixels) as side information (See Fig. 3.6). Both images are partitioned into non-overlapping patches, where vectors $\mathbf{x}_1$ represent 8×8 patches extracted from the signal of interest, and vectors $\mathbf{x}_2$ represent 2×2 patches extracted from the lower resolution version of the same image, that are partitioned into overall 4096 patches for both images. We assume that the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a joint GMM with $K_1 = K_2 = 20$. The parameters of the joint GMM are obtained via the EM algorithm, learned from from the Caltech 101 dataset [97] (as in previous experiment). In this experiment, we also consider both designed cases when side information is provided at the decoder and at the encoder and the decoder (suboptimal design as in previous experiment), as well as the random case when the projection kernel is generated with i.i.d., zero-mean, Gaussian entries with fixed variance.

More reconstruction results with different number of linear measurements when $m_1 = 5$ and $m_1 = 10$ and with noise level $\sigma^2 = -40$ dB, are reported in Figs. 3.8 and 3.9, respectively. From the top to the bottom, the reconstruction images correspond the following different scenarios: (a) random projection kernel with i.i.d., zero-mean, Gaussian entries with fixed variance with side information at the decoder only; (b) designed projection kernel with side information at the decoder only; (c) designed kernel with side information at both the encoder and the decoder, i.e., we consider the suboptimal design approach based on the classification of vectors $\mathbf{y}_1$ and $\mathbf{x}_2$ using MAP classifier (as in previous experiment).

We can observe that measurement designs in Figs. 3.8(b) and (c) can lead to PSNR gains of approximately 3 dB, in comparison to the case of random projection kernel in Fig. 3.8(a). On the other hand, the reconstruction results when $m_1 = 10$ illustrate that measurement designs in Figs. 3.9(b) and (c) exhibit PSNR gains of approximately 5 dB with respect to the case of random projection kernels in Fig. 3.9(a). The suboptimal design based on the classification of vectors $\mathbf{y}_1$ and $\mathbf{x}_2$ (see Figs. 3.8(c) and 3.9(c)), yields PSNR values slightly lower than those obtained with side information at the decoder only as shown in Figs. 3.8(b) and 3.9(b). Moreover, we also observe that using a high number of projections in designing projection kernel can apparently improve the reconstruction performance than the lower one.

Moreover, we also provide the reconstruction PSNR results with different number of noise levels with different number of measurements $m_1 = 5, 10, 15$ as reported in Table 3.1. It is clear to see the reconstruction PSNR are slightly improved with increasingly number of measurements and dropping noise levels. We can also observe that the designed cases (b) and (c) can lead to significant PSNR gains with respect to the random measurement kernels. In particular, the suboptimal design approach achieves similar PSNR values as when the projection kernel is designed via more computationally expensive numerical solution of the optimization problem.

Table 3.1: PSNR values (dB) of the reconstruction with $K_1 = K_2 = 20$ with number of measurements $m_1 = 5, 10, 15$.

Noise Level	-20 dB	-30 dB	-40 dB	-50 dB	-60 dB
	$m_1 = 5$				
case (a)	27.8805 dB	27.8871 dB	28.0517 dB	28.6590 dB	29.3019 dB
case (b)	29.7853 dB	31.0945 dB	31.6349 dB	31.7183 dB	31.7589 dB
case (c)	29.7068 dB	30.0012 dB	30.2888 dB	30.3212 dB	30.3245 dB
	$m_1 = 10$				
case (a)	27.8801 dB	27.8993 dB	28.0611 dB	28.9214 dB	30.3993 dB
case (b)	29.8251 dB	33.5534 dB	33.6339 dB	33.9100 dB	34.2226 dB
case (c)	29.7790 dB	32.3125 dB	33.5103 dB	33.6654 dB	33.6825 dB
	$m_1 = 15$				
case (a)	27.8809 dB	27.9074 dB	28.1607 dB	28.9373 dB	30.7489 dB
case (b)	30.1744 dB	32.8475 dB	35.9751 dB	36.2720 dB	36.3205 dB
case (c)	30.0270 dB	33.3483 dB	35.6850 dB	36.0598 dB	36.1072 dB

(a) Random kernel, side information at decoder

(b) Designed kernel, side information at decoder

(c) Designed kernel, side information at encoder and decoder

Figure 3.7: Reconstruction results of the image "Lena" for $\sigma^2 = -40$ dB and measurements $m_1 = 15$, from the top to the bottom, the reconstruction PSNR values are 28.1607 dB, 35.9751 dB, 35.6850 dB.

(a) Random kernel, side information at decoder

(b) Designed kernel, side information at decoder

(c) Designed kernel, side information at encoder and decoder

Figure 3.8: Reconstruction results of the image "Lena" for $\sigma^2 = -40$ dB and measurements $m_1 = 5$, from the top to the bottom, the reconstruction PSNR values are 28.0517 dB, 31.6349 dB, 30.2888 dB.

(a) Random kernel, side information at decoder

(b) Designed kernel, side information at decoder

(c) Designed kernel, side information at encoder and decoder

Figure 3.9: Reconstruction results of the image "Lena" for $\sigma^2 = -40$ dB and measurements $m_1 = 10$, from the top to the bottom, the reconstruction PSNR values are 28.0611 dB, 33.6339 dB, 33.5103 dB.

We now study the case by considering the signal of interest and the side information are divided into larger size patches. We also use a high-resolution image "Lena" with resolution 512×512 as the signal of interest and a low-resolution version of the same subject (128×128 pixels) as side information (See Fig. 3.6), by assuming both the signal of interest and the side information are divided into 1024 patches, i.e., the vectors $\mathbf{x}_1$ extracted form the high-resolution image are divided with 16×16 non-overlapping patches and vectors $\mathbf{x}_2$ extracted form the low-resolution image are divided with 4×4 non-overlapping patches, respectively. In this experiment, we consider both vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint GMM prior with $K_1 = K_2 = 20$ classes. All joint GMM parameters are trained by the EM algorithm over a set of patches obtained from images in the Caltech 101 dataset. We also consider a projection kernel $\mathbf{\Phi}_1$ that is randomly obtained with i.i.d., zero-mean, Gaussian entries with fixed variance, as comparison purpose.

Fig. 3.10 shows the reconstruction results with PSNR values with measurements $m_1 = 15$ and $\sigma^2 = -40$ dB. From the top to the bottom, the reconstruction images correspond the following different scenarios: (a) random projection kernel with i.i.d., zero-mean, Gaussian entries with fixed variance with side information at the decoder only; (b) designed projection kernel with side information at the decoder only; (c) designed kernel with side information at both the encoder and the decoder, i.e., we consider the suboptimal design approach based on the classification of vectors $\mathbf{y}_1$ and $\mathbf{x}_2$ using MAP classifier (as in previous experiments).

Both designed cases in Figs. 3.10(b) and (c) provide PSNR gains of approximately 4 dB, in relation to the random kernel case as shown in Fig. 3.10(a). We can also observe that patches extracted with larger size from both the signal of interest and the side information does not improve the reconstruction performance, exhibiting PSNR values lower than those obtained with lower dimensional patches. This result implies that the GMM is not rich enough to describe such higher patches well. Namely, we need to increase the number of GMM components to model higher dimensional patches.

For a higher dimensional patches, we also consider the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a joint GMM distribution with $K_1 = K_2 = 30$ classes. Fig. 3.11 shows the reconstruction images with corresponding PSNR values, with the fixed number of linear measurements $m_1 = 15$ and with noise level $\sigma^2 = -40$ dB. The results in Fig. 3.11 also revel that both designs can improve the reconstruction performance with respect to random projection kernels. We can also observe that the patches described by a 30-classes joint GMM distribution yields PSNR values higher than those obtained with 20-classes joint GMM distribution.

(a) Random kernel, side information at decoder

(b) Designed kernel, side information at decoder

(c) Designed kernel, side information at encoder and decoder

Figure 3.10: Reconstruction results of the image "Lena" for $\sigma^2 = -40$ dB, measurements $m_1 = 15$, and with $K_1 = K_2 = 20$, from the top to the bottom, the reconstruction PSNR values are 27.1351 dB, 31.3136 dB, 31.0966 dB.

(a) Random kernel, side information at decoder

(b) Designed kernel, side information at decoder

(c) Designed kernel, side information at encoder and decoder

Figure 3.11: Reconstruction results of the image “Lena” for $\sigma^2 = -40$ dB, measurements $m_1 = 15$, and with $K_1 = K_2 = 30$, from the top to the bottom, the reconstruction PSNR values are 27.8558 dB, 32.2039 dB, 32.1664 dB.

3.5 Summary

This chapter illustrates the impact of projection kernel design on the reconstruction performance in the presence of side information when both input signal and the side information are jointly drawn from both Gaussian and GMM distributions, with possibly low-rank, class-conditioned covariance matrices. In particular, we have considered both cases of kernel design when side information is available at decoder only, and at the encoder and the decoder. The case for random projection kernel is also used for comparison purposes.

We have provided sharp necessary and sufficient conditions on the number of linear projections that guarantee the reconstruction MMSE to approach zero in the low-noise regime, for both cases when the side information is available at the decoder or at the decoder and encoder. We have observed that the case when side information is available at decoder only, and at both the encoder and the decoder exhibit the same characterization of the minimum number of measurements needed for the MMSE phase transition.

We have also provided numerical experiments with synthetic data for both Gaussian and GMM signals to confirm our proposed theorems align well with proposed theorems. The results have revealed that even though measurement design has no impact on phase transition, careful projection kernel designs can lead to significant reduction of the reconstruction error for finite noise levels. We have also observed that kernel design schemes that leverage the presence of side information at both the encoder and the decoder do not provide significant advantages with respect to kernel designs with side information at the decoder only. Moreover, numerical results are also presented to showcase the impact of projection design with real imaging data with side information.

Appendices

Appendix A

Proofs and Derivations of Chapters 3

A.1 Proof of Theorem 22 and 23

The sufficient conditions in (3.13) and (3.14) can be obtained by analyzing upper bounds corresponding to the MMSE associated with designed kernels with the MMSE obtained with random projection kernels and by using the results in [81, Theorem 4].

We now consider the necessary conditions for both designs in (3.13) and (3.14). On adopting the notation $\text{MMSE}(\mathbf{u}|\mathbf{v}) = \mathbb{E}[\|\mathbf{u} - \mathbb{E}[\mathbf{u}|\mathbf{v}]\|^2]$, we can write the MMSE with side information as

$$\text{MMSE}(\sigma^2, \mathbf{\Phi}_1) = \mathbb{E}[\text{MMSE}(\sigma^2, \mathbf{\Phi}_1, \mathbf{x}_2)] \tag{A.1}$$

$$= \mathbb{E}[\text{MMSE}(\mathbf{u}|\mathbf{v})], \tag{A.2}$$

where the expectation is taken with respect to $\mathbf{x}_2$, and where $\mathbf{u} \sim p(\mathbf{x}_1|\mathbf{x}_2)$ and $\mathbf{v} = \mathbf{\Phi}_1\mathbf{u} + \mathbf{w}_1$. In particular, the reconstruction MMSE does not depend on the value of the mean of the input signal, we then have $\text{MMSE}(\sigma^2, \mathbf{\Phi}_1) = \text{MMSE}(\mathbf{u}|\mathbf{v})$. We also note that the vector $\mathbf{u}$ is drawn from a Gaussian distribution with class-conditioned means and covariance matrices, given by

$$\boldsymbol{\mu}_{\mathbf{u}} = \boldsymbol{\mu}_{\mathbf{x}_1} + \mathbf{\Sigma}_{\mathbf{x}_{12}}(\mathbf{\Sigma}_{\mathbf{x}_2})^{\dagger}(\mathbf{x}_2 - \boldsymbol{\mu}_{\mathbf{x}_2}), \tag{A.3}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{u}} = \boldsymbol{\Sigma}_{\mathbf{x}_1} - \boldsymbol{\Sigma}_{\mathbf{x}_{12}}(\boldsymbol{\Sigma}_{\mathbf{x}_2})^{\dagger}\boldsymbol{\Sigma}_{\mathbf{x}_{21}}, \tag{A.4}$$

where the symbol $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse [98]. Then, by leveraging [93, Theorem 3], we have that a necessary condition for $\mathrm{MMSE}(\mathbf{u}|\mathbf{v})$ to approach zero in the low-noise regime is given by

$$m_1 \geq \mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{u}}). \tag{A.5}$$

Such necessary condition holds also when $\boldsymbol{\Phi}_1$ is designed in order to minimize the MMSE, thus implying that the necessary condition for the phase transition of $\mathrm{MMSE}(\mathbf{u}|\mathbf{v})$ holds for both cases when side information is available at the decoder and at both the encoder and the decoder (i.e., when the designed $\boldsymbol{\Phi}_1$ is a function of the current realization of $\mathbf{x}_2$).

By using a result on the generalized Schur complement $\boldsymbol{\Sigma}_{\mathbf{x}_2}$ of a positive semidefinite matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ [99], we finally obtain

$$\mathrm{rank}(\boldsymbol{\Sigma}_{\mathbf{u}}) = \ r_{\mathbf{x}} - r_{\mathbf{x_2}}, \tag{A.6}$$

A.2 Proof of Theorem 24 and 25

The sufficient conditions (3.16) and (3.18) can be obtained by upper bounding the MMSE associated with designed kernels with the MMSE obtained with random projection kernels and by using the results in [81, Theorem 4].

Consider now the necessary conditions (3.15) and (3.17). On adopting the notation $\mathrm{MMSE}(\mathbf{u}|\mathbf{v})$ in (A.1) and (A.2), the expectation is taken with respect to $\mathbf{x}_2$, and where $\mathbf{u} \sim p(\mathbf{x}_1|\mathbf{x}_2)$ and $\mathbf{v} = \boldsymbol{\Phi}_1\mathbf{u} + \mathbf{w}_1$. We note that the vector $\mathbf{u}$ is distributed according to a GMM with class-conditioned means and covariance matrices. For a given value $\mathbf{x}_2$, it holds

$$p(\mathbf{u}) = \sum_{i=1}^{K_1}\sum_{k=1}^{K_2} p_{C_1,C_2}(i,k)p(\mathbf{x}_1|\mathbf{x}_2, C_1 = i, C_2 = k) \tag{A.7}$$

$$= \sum_{i=1}^{K_1}\sum_{k=1}^{K_2} p_{C_1,C_2}(i,k)\frac{p(\mathbf{x}_1, \mathbf{x}_2, C_1 = i, C_2 = k)}{p(\mathbf{x}_2, C_1 = i, C_2 = k)} \tag{A.8}$$

$$= \sum_{i=1}^{K_1}\sum_{k=1}^{K_2} \tilde{p}_{C_1,C_2}(i,k)\mathcal{N}(\boldsymbol{\mu}_{\mathbf{u}}^{(i,k)}, \boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)}) \tag{A.9}$$

Then we obtain

$$\boldsymbol{\mu}_{\mathbf{u}}^{(i,k)} = \boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)} + \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})^{\dagger}(\mathbf{x}_2 - \boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)}) \tag{A.10}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)} = \boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})^{\dagger}\boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}, \tag{A.11}$$

where the symbol $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse [98]. Then, by leveraging [93, Theorem 6], we have that a necessary condition for MMSE($\mathbf{u}|\mathbf{v}$) to approach zero in the low-noise regime is given by

$$m_1 \geq \max_{(i,k)} \operatorname{rank}(\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)}). \tag{A.12}$$

Note that such necessary condition holds also when $\boldsymbol{\Phi}_1$ is designed in order to minimize the MMSE, thus implying that the necessary condition for the phase transition of MMSE($\mathbf{u}|\mathbf{v}$) holds for both cases when side information is available at the decoder and at both the encoder and the decoder (i.e., when the designed $\boldsymbol{\Phi}_1$ is a function of the current realization of $\mathbf{x}_2$). Finally, we note that $\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)}$ is the generalized Schur complement of $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ of the positive semidefinite matrix $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$. Then, by using the result in [99], we have

$$\operatorname{rank}(\boldsymbol{\Sigma}_{\mathbf{u}}^{(i,k)}) = r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x_2}}^{(i,k)}, \tag{A.13}$$

which concludes the proof.

Chapter 4

Measurement Design with Side Information – Capturing the Side Information

In this chapter, we study the problem of how to optimally capture side information to aid in the reconstruction of high-dimensional signals from low-dimensional measurements. We also consider both the signal of interest and the side information are jointly described by either the Gaussian or GMM distributions.

We then focus on deriving the sufficient and (occasionally) necessary conditions on the number of measurements needed to guarantee that reconstruction error tends to be zero in the low-noise regime for both Gaussian and GMM sources. In particular, we consider the case where the projection kernel associated to the signal of interest is random, whereas the projection kernel associated to the side information is designed. We also consider the scenario where both projection kernels are randomly constituted for comparison purposes. Moreover, we derive closed-form linear side information measurement designs for the reconstruction MMSE to approach zero in the low-noise regime.

Finally, we provide a number of numerical results both with synthetic data for Gaussian and GMM signals, in order to confirm that our analysis is aligned very well with practice. Numerical results are also presented to showcase the impact of projection design on applications with real imaging data. At the end of this chapter, we conclude with the main contributions.

4.1 Problem Statement

4.1.1 Model

Fig. 4.1 depicts the compressive measurement model under consideration where $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ is the signal of interest and the $\mathbf{x}_2 \in \mathbb{R}^{n_2}$ is the side information. Our objective is to reconstruct the signal of interest both from the linear measurements $\mathbf{y}_1 \in \mathbb{R}^{m_1}$, with $n_1 \geq m_1$, that are given by

$$\mathbf{y}_1 = \mathbf{\Phi}_1 \mathbf{x}_1 + \mathbf{w}_1; \tag{4.1}$$

as well as the noisy, linear, compressive measurements $\mathbf{y}_2 \in \mathbb{R}^{m_2}$, with $n_2 \geq m_2$, given by

$$\mathbf{y}_2 = \mathbf{\Phi}_2 \mathbf{x}_2 + \mathbf{w}_2. \tag{4.2}$$

The matrices $\mathbf{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ and $\mathbf{\Phi}_2 \in \mathbb{R}^{m_2 \times n_2}$ represent linear projection kernels and $\mathbf{w}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma_1^2)$ and $\mathbf{w}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \cdot \sigma_2^2)$ are additive white Gaussian noise (AWGN) vectors that model possible distortion introduced by the sensing process and noise.

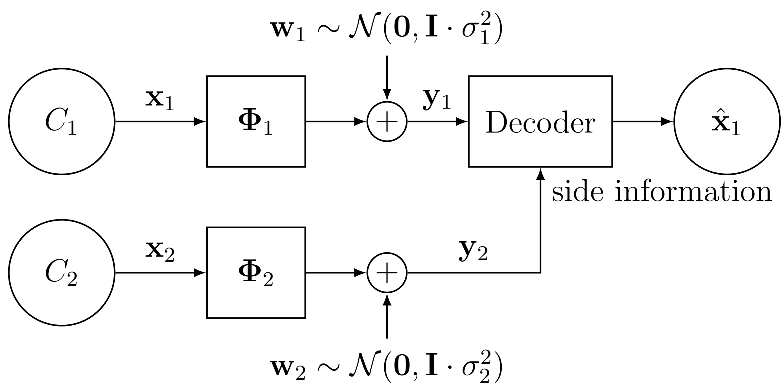


Figure 4.1: Compressive sensing model in the presence of side information.

We are also interested in the reconstruction of the signal of interest $\mathbf{x}_1$ estimated via the noisy linear measurements $\mathbf{y}_1$ and $\mathbf{y}_2$ using the optimal (in MMSE sense) conditional estimator given by

$$\hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}[\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2] = \int_{-\infty}^{+\infty} \mathbf{x}_1 p(\mathbf{x}_1 | \mathbf{y}_1, \mathbf{y}_2) d\mathbf{x}_1, \tag{4.3}$$

where $p(\mathbf{x}_1|\mathbf{y}_1, \mathbf{y}_2)$ is the *a posteriori* pdf of $\mathbf{x}_1$ given the measurements $\mathbf{y}_1$ and $\mathbf{y}_2$. Therefore, the reconstruction performance is measured in terms of the MMSE which is given by:

$$\mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = \mathbb{E}[||\mathbf{x}_1 - \hat{\mathbf{x}}_1(\mathbf{y}_1, \mathbf{y}_2)||^2], \tag{4.4}$$

Note that we express the MMSE explicitly as a function of the noise variances σ_1^2, σ_2^2 and projection kernels $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$.

We will be assuming that the signal of interest $\mathbf{x}_1$ and the side information signal $\mathbf{x}_2$ – which are correlated – are described by a joint GMM, characterized by underlying class labels $C_1 \in \{1, \dots, K_1\}$ and $C_2 \in \{1, \dots, K_2\}$, respectively, which obey the joint pmf $P_{C_1,C_2}(i, k)$. Conditioned on the underlying class labels $C_1 = i$ and $C_2 = k$, the joint distribution of $\mathbf{x}_1$ and $\mathbf{x}_2$ follow the joint Gaussian pdf:

$$p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}), \tag{4.5}$$

so that $\mathbf{x}_1$ and $\mathbf{x}_2$ obey the GMM given by

$$\begin{aligned} p(\mathbf{x}_1, \mathbf{x}_2) &= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1,C_2}(i, k) p(\mathbf{x}_1, \mathbf{x}_2 | C_1 = i, C_2 = k) \\ &= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} P_{C_1,C_2}(i, k) \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}, \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}), \end{aligned} \tag{4.6}$$

where

$$\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)} \\ \boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}, \quad \mathbf{\Sigma}_{\mathbf{x}}^{(i,k)} = \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)} & \mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \\ \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} & \mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} \end{bmatrix}. \tag{4.7}$$

In other terms, conditioned on a pair of classes $C_1 = i$ and $C_2 = k$, $\mathbf{x}_1$ is Gaussian distributed with mean $\boldsymbol{\mu}_{\mathbf{x}_1}^{(i,k)}$ and covariance matrix $\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)}$, and likewise $\mathbf{x}_2$ is Gaussian distributed with mean $\boldsymbol{\mu}_{\mathbf{x}_2}^{(i,k)}$ and covariance matrix $\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}$, respectively. In addition, $\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}$ represents the cross-covariance between $\mathbf{x}_1$ and $\mathbf{x}_2$ given the classes $C_1 = i$ and $C_2 = k$. Here, $r_{\mathbf{x}}^{(i,k)} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}) \le n_1 + n_2$, $r_{\mathbf{x}_1}^{(i,k)} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)}) \le n_1$ and $r_{\mathbf{x}_2}^{(i,k)} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}) \le n_2$. Note that, conditioned on class labels $C_1 = i, C_2 = k$, such ranks can also act as proxy to measure the correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$. In fact, for all (i, k), the overall rank $r_{\mathbf{x}}^{(i,k)}$ is upper bounded by

$r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)}$ and lower bounded by $\max\{r_{\mathbf{x}_1}^{(i,k)}, r_{\mathbf{x}_2}^{(i,k)}\}$; the upper bound is attained by uncorrelated signals (conditioned on class labels $C_1 = i, C_2 = k$) and the lower bound is attained by perfectly correlated ones (conditioned on class labels $C_1 = i, C_2 = k$).

As in Chapter 3, we will also consider that the linear spaces associated to the images of the covariance matrices $\mathbf{\Sigma}_{\mathbf{x}}^{(i,k)}$, $\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)}$ and $\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ for different class labels $C_1 = i$ and $C_2 = k$ are independently drawn at random from a continuous pdf over the corresponding Grassmann manifold.

4.2 Design Method

We now introduce our measurement design problem. We consider a scenario where the projection matrix associated to the signal of interest is random, whereas the projection matrix associated to the side information is designed. We also consider the scenario where both projection matrices are randomly constituted[1] for comparison purposes.

In particular, the designed measurement matrix $\mathbf{\Phi}_2^\star$ corresponds to the solution of the optimization problem:

$$\begin{aligned} &\underset{\mathbf{\Phi}_2}{\text{minimize}}\ \mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) \\ &\text{subject to } \mathrm{tr}(\mathbf{\Phi}_2\mathbf{\Phi}_2^{\mathrm{T}}) \le m_2, \end{aligned} \tag{4.8}$$

where the trace constraint in (4.8) limits the average energy associated to the projection kernel corresponding to the side information.

We also note that the decoder is assumed to have knowledge of the joint pdf $p(\mathbf{x}_1, \mathbf{x}_2)$, the linear projection kernels $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, and the noise variances $\boldsymbol{\sigma}_1^2$ and $\boldsymbol{\sigma}_2^2$.

Our focus, then, is to provide sufficient and (occasionally) necessary conditions on the number of measurement taken both from the signal of interest and the side information, to guarantee that the MMSE tends to be zero in the low-noise regime, i.e.

$$\lim_{\sigma_1^2, \sigma_2^2 \to 0} \mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 0, \tag{4.9}$$

[1]Note that the random measurement matrices $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are drawn from left-rotationally invariant distributions as in Chapter 2.

Such conditions will be shown both for the scenario where the measurement matrices $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are randomly constituted, and for the scenario where measurement matrix $\mathbf{\Phi}_1$ is random but the measurement matrix associated to the side information is designed when $\mathbf{\Phi}_2 = \mathbf{\Phi}_2^\star$.

The challenge associated with the characterization of sufficient and (occasionally) necessary conditions on the number of measurements for reliable reconstruction (i.e., such that (4.9) holds) is due to the fact that the (4.4) does not admit a closed form expression, even though (4.3) does for GMMs (the problem is also compounded in view of the fact that the solution to (4.8) cannot be analytically characterized). Therefore, our ensuing analysis will rely on bounds to the MMSE as a means to characterize conditions for the MMSE to approach zero in the low-noise regime.

4.3 Bounds on the Number of Measurements for Reliable Reconstruction

We focus now on determining sufficient conditions (and necessary conditions in some cases) on the minimum number of measurements m_1 and m_2 for reliable reconstruction (i.e., such that (4.9) holds), assuming that the measurement matrix associated to the signal of interest is randomly constituted, and the measurement matrix associated to the side information is optimally designed by solving the optimization problem of (4.8). We consider both the simpler scenario where the signal of interest and the side information obey a joint multivariate Gaussian distribution as well as the more general scenario where the signals obey a joint GMM.

4.3.1 Gaussian Sources

We start by considering the case where $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a joint Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{x}}$ and covariance matrix $\mathbf{\Sigma}_{\mathbf{x}}$, i.e., when $K_1 = K_2 = 1$ in (4.6).

The following theorem considers the scenario where the matrix used to capture the side information is optimally designed.

Theorem 26. *Consider the measurements model in (4.1) and (4.2), where* $\mathbf{x}_1$ *and*

$\mathbf{x}_2$ *are described by a joint Gaussian distribution with mean* $\boldsymbol{\mu}_{\mathbf{x}}$ *and covariance* $\boldsymbol{\Sigma}_{\mathbf{x}}$*, such that* $r_{\mathbf{x}} = rank(\boldsymbol{\Sigma}_{\mathbf{x}})$*,* $r_{\mathbf{x}_1} = rank(\boldsymbol{\Sigma}_{\mathbf{x}_1})$ *and* $r_{\mathbf{x}_2} = rank(\boldsymbol{\Sigma}_{\mathbf{x}_2})$*. Assume that* $\boldsymbol{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ *is drawn from a left-rotationally invariant distribution, and* $\boldsymbol{\Phi}_2 = \boldsymbol{\Phi}_2^\star$ *is the optimal projection matrix corresponding to the solution of the optimization problem in* (4.8)*. Then, it holds*

$$\lim_{\sigma_1^2,\sigma_2^2 \to 0} MMSE(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2^\star) = 0 \Leftrightarrow \begin{cases} m_1 + m_2 \geq r_{\mathbf{x}_1} \\ m_1 \geq r_{\mathbf{x}} - r_{\mathbf{x}_2} \end{cases} . \tag{4.10}$$

Proof: See Appendix B.1. ■

The necessary and sufficient conditions for reliable reconstruction in Theorem 26 suggest that:

1. The total number of measurements $m_1 + m_2$ must be equal to at least the dimension of the space spanned by the signal of interest $r_{\mathbf{x}_1}$. We recall that, without side information we would need $m_1 \geq r_{\mathbf{x}_1}$, which means that the m_2 measurements taken from the side information signal allow to decrease the number of measurements m_1 required for reliable reconstruction by m_2. Note that this does not happen in general for the case when side information measurements are random, since, in that case, the number of measurements $m_1 + m_2$ must be equal to at least $r_{\mathbf{x}} \geq r_{\mathbf{x}_1}$ in Theorem 20.

2. The number of measurements taken from $\mathbf{x}_1$ must also be at least equal to the codimension of the intersection of the space spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$ in the space spanned by $\mathbf{x}_1$. In other terms, this means that the number of measurements taken from $\mathbf{x}_1$ must be large enough to capture the specific attributes of $\mathbf{x}_1$ that cannot be inferred from the observation of $\mathbf{x}_2$.

The results in Theorem 26 offers a sharp characterization of necessary and sufficient conditions on the number of measurements needed for reliable reconstruction for Gaussian sources. We also describe a constructive method to generate a projection kernel $\boldsymbol{\Phi}_2$ that allows reliable reconstruction with the minimum number of measurements predicted by Theorem 26 (see Appendix). Such design is obtained from the generalized singular value decomposition (GSVD) [100] associated with two matrices related to the covariance matrices of the sources, $\mathbf{A} = (\mathbf{\Sigma}_{\mathbf{x}_2})^{1/2}$ and $\mathbf{B} = (\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}})^{1/2}$. In particular, these matrices can be written as follows:

$$\mathbf{A} = \mathbf{U}\mathbf{C}\mathbf{X}^{\mathrm{T}} \quad , \quad \mathbf{B} = \mathbf{V}\mathbf{S}\mathbf{X}^{\mathrm{T}}, \tag{4.11}$$

where $\mathbf{U} \in \mathbb{R}^{n_2 \times n_2}$, $\mathbf{V} \in \mathbb{R}^{n_2 \times n_2}$ are unitary matrices, $\mathbf{X} \in \mathbb{R}^{n_2 \times n_2}$ is non-singular, and $\mathbf{C} = [\mathbf{\Lambda_A}\, \mathbf{0}]$ and $\mathbf{S} = [\mathbf{\Lambda_B}\, \mathbf{0}]$ are diagonal matrices with

$$\mathbf{\Lambda_A} = \begin{array}{c c} & \begin{array}{cc} r_{\mathbf{x}_1}+r_{\mathbf{x}_2}-r_{\mathbf{x}} & r_{\mathbf{x}}-r_{\mathbf{x}_1} \end{array} \\ \begin{array}{c} r_{\mathbf{x}_1}+r_{\mathbf{x}_2}-r_{\mathbf{x}} \\ r_{\mathbf{x}}-r_{\mathbf{x}_1} \end{array} & \begin{bmatrix} \mathbf{I} & \\ & \mathbf{D_A} \end{bmatrix} \end{array}, \tag{4.12}$$

$$\mathbf{\Lambda_B} = \begin{array}{c c} & \begin{array}{cc} r_{\mathbf{x}_1}+r_{\mathbf{x}_2}-r_{\mathbf{x}} & r_{\mathbf{x}}-r_{\mathbf{x}_1} \end{array} \\ \begin{array}{c} r_{\mathbf{x}_1}+r_{\mathbf{x}_2}-r_{\mathbf{x}} \\ r_{\mathbf{x}}-r_{\mathbf{x}_1} \end{array} & \begin{bmatrix} \mathbf{0} & \\ & \mathbf{D_B} \end{bmatrix} \end{array}. \tag{4.13}$$

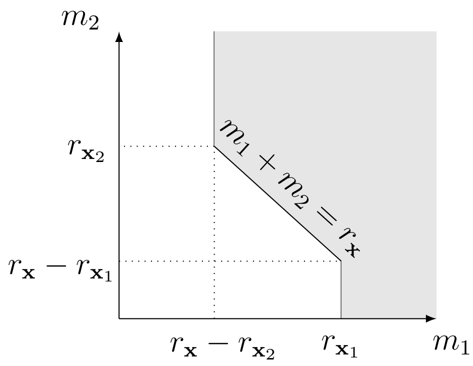
where we have reported explicitly the dimensions of the square blocks in $\mathbf{\Lambda_A}$ and $\mathbf{\Lambda_B}$.

Then, the proposed design is given by

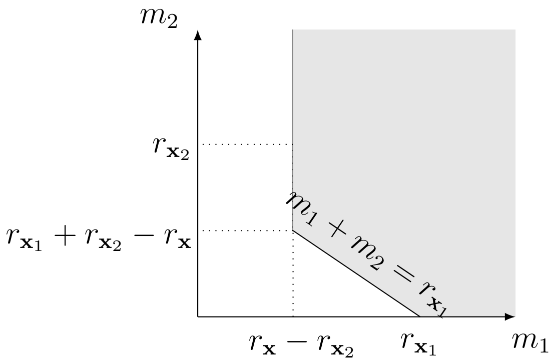
$$\mathbf{\Phi}_2 = \begin{bmatrix} \mathbf{I}_{m_2'} & \mathbf{0}_{m_2' \times (n_2 - m_2')} \\ \mathbf{0}_{(m_2 - m_2') \times m_2'} & \mathbf{0}_{(m_2 - m_2') \times (n_2 - m_2')} \end{bmatrix} \mathbf{X}^{-1}. \tag{4.14}$$

where $m_2' = \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}$.

This shows – in view of the sharpness of the necessary and sufficient conditions in (4.10) – that an optimal matrix design exhibits the following attribute: it captures the portion of the linear space contained in the intersection between the image of $\mathbf{\Sigma}_{\mathbf{x}_2}$ and the null space of the matrix $\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}$. Then, noting that the conditional distribution $p(\mathbf{x}_2|\mathbf{x}_1)$ is Gaussian and with covariance matrix given by $\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}$, we can observe that an optimal matrix design is aligned with the linear space spanned by signals drawn from the distribution $p(\mathbf{x}_2)$ which is not occupied by signals drawn from the conditional distribution $p(\mathbf{x}_2|\mathbf{x}_1)$ in order to improve the reconstruction performance. Intuitively, such



(a) Random projection kernel.



(b) Designed projection kernel.

Figure 4.2: Representation of the conditions on m_1 and m_2 for MMSE phase transition for Gaussian sources, for the random kernel case (a) and designed kernel case (b).

design aims to capture the most of signal components from $\mathbf{x}_2$ that are maximally correlated with $\mathbf{x}_1$.

The result in Theorem 26 together with the result in Theorem 20 also showcase the merit of an optimal measurement matrix $\mathbf{\Phi}_2^\star$ in comparison to a random one, in terms of the number of measurements necessary and sufficient for reliable reconstruction.

Corollary 27. *Consider the measurement and source model in Theorem 26. Let $\mathcal{M}_{\mathsf{G}}^{\mathsf{R}}$ be the set of pairs (m_1, m_2) such that $\lim_{\sigma_1^2,\sigma_2^2 \to 0} \mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 0$ for random $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, and let $\mathcal{M}_{\mathsf{G}}^{\mathsf{D}}$ be the set of pairs (m_1, m_2) such that $\lim_{\sigma_1^2,\sigma_2^2 \to 0} \mathrm{MMSE}(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2^\star) = 0$ for random $\mathbf{\Phi}_1$ and the optimal $\mathbf{\Phi}_2^\star$. Then, it holds*

$$\mathcal{M}_{\mathsf{G}}^{\mathsf{R}} \subseteq \mathcal{M}_{\mathsf{G}}^{\mathsf{D}}. \tag{4.15}$$

Proof: See Appendix B.2. ■

The advantage of using a designed matrix in relation to a random one to capture side information is also illustrated in Fig. 4.2. The shaded region of Fig. 4.2(a) represents the set $\mathcal{M}_{\mathsf{G}}^{\mathsf{R}}$, as reported in Theorem 20; and, Fig. 4.2(b) represents the set $\mathcal{M}_{\mathsf{G}}^{\mathsf{D}}$.

We can note that, depending on the specific values of $r_{\mathbf{x}}, r_{\mathbf{x}_1}, r_{\mathbf{x}_2}$, i.e., depending on the correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$, proper design of the linear projection kernel $\mathbf{\Phi}_2$ can guarantee a significant advantage with respect to random kernels in terms of number of measurements required for reliable reconstruction. This is somewhat surprising because it has been shown that optimal design of the linear projection kernel used to acquire the signal of interest does not reduce significantly the minimum number of measurements required for reliable reconstruction in Theorems 22 and 23 in Chapter 3.

This fact can be justified by the following observation. In the case of Gaussian sources, the MMSE associated to the estimation of $\mathbf{x}_1$ from the observation of $\mathbf{y}_1$ and $\mathbf{y}_2$ is equivalent to the estimation of the vector $\mathbf{z} \sim p(\mathbf{x}_1|\mathbf{y}_2)$ from the observation of the vector $\mathbf{\Phi}_1\mathbf{z} + \mathbf{w}_1$. Moreover, as it was shown in [93], the fundamental limits on the number of measurements needed to verify (4.9) are dictated by the dimension of the linear subspace spanned by the signal of interest, for both cases when $\mathbf{\Phi}_1$ is random and designed. Therefore, careful design of $\mathbf{\Phi}_1$ does not have an impact on such limit, whereas careful design of $\mathbf{\Phi}_2$ can reduce the dimension of the space spanned by signals $\mathbf{z} \sim p(\mathbf{x}_1|\mathbf{y}_2)$ in the low-noise regime, thus leading to a reduction of the minimum number of measurements needed for reliable reconstruction.

4.3.2 GMM Sources

We now consider the more challenging scenario where $\mathbf{x}_1$ and $\mathbf{x}_2$ obey a joint GMM. In this case, we only provide sufficient conditions rather than sufficient and necessary conditions for reliable reconstruction.

The following theorem considers the scenario where the matrix used to capture the side information is optimally designed.

Theorem 28. *Consider the measurements model in (4.1) and (4.2), where* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *conditioned on the underlying class labels* $C_1 = i$, $C_2 = k$ *obey a joint Gaussian* *distribution with mean* $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)}$ *and covariance* $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$ *such that* $r_{\mathbf{x}}^{(i,k)} = rank(\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)})$, $r_{\mathbf{x}_1}^{(i,k)} = rank(\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})$ *and* $r_{\mathbf{x}_2}^{(i,k)} = rank(\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})$ $\forall i, k$. *Consider also that the linear spaces associated to the images of the covariance matrices* $\boldsymbol{\Sigma}_{\mathbf{x}}^{(i,k)}$, $\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)}$ *and* $\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)}$ *are independently drawn from a continuous pdf over the corresponding Grassmann manifold. Assume that* $\boldsymbol{\Phi}_1 \in \mathbb{R}^{m_1 \times n_1}$ *is random, drawn from a left-rotationally invariant distribution, and* $\boldsymbol{\Phi}_2 = \boldsymbol{\Phi}_2^\star$ *is the optimal projection matrix corresponding to the solution of the optimization problem in* (4.8), *sufficient conditions on the number of projections* m_1 *and* m_2 *for*

$$\lim_{\sigma_1^2, \sigma_2^2 \to 0} MMSE(\sigma_1^2, \sigma_2^2, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2^\star) = 0 \tag{4.16}$$

are given by

$$m_1 > \begin{cases} r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)} & , \text{if } m_2 \leq r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} \\ \min\{r_{\mathbf{x}}^{(i,k)} - m_2, r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)}\} & , \text{if } r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \leq r_{\mathbf{x}_2}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} & , \text{if } m_2 > r_{\mathbf{x}_2}^{(i,k)} \end{cases} \tag{4.17}$$

for $i = 1, \ldots K_1$ *and* $k = 1, \ldots, K_2$, *where* $m_2^{(i,k)} \in \mathbb{N}$ *are such that* $m_2^{(i,k)} \leq r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}$ *and* $\sum_{i,k} m_2^{(i,k)} = m_2$.

Proof: See Appendix B.3. ∎

The sufficient conditions for reliable reconstruction for GMM sources embodied in Theorem 28 are obtained by considering a specific (suboptimal) design for $\boldsymbol{\Phi}_2$, which is inspired by the projection design which achieves the necessary and sufficient conditions for reliable reconstruction for Gaussian sources in Theorem 26. This suboptimal design is obtained from the GSVD associated with the pair of matrices

$$\mathbf{A}^{(i,k)} = (\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)})^{1/2} \tag{4.18}$$

$$\mathbf{B}^{(i,k)} = (\boldsymbol{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \boldsymbol{\Sigma}_{\mathbf{x}_{21}}^{(i,k)} (\boldsymbol{\Sigma}_{\mathbf{x}_1}^{(i,k)})^\dagger \boldsymbol{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})^{1/2}, \tag{4.19}$$

for all (i,k). In particular, we write these matrices as follows:

$$\mathbf{A}^{(i,k)} = \mathbf{U}^{(i,k)}\mathbf{C}^{(i,k)}(\mathbf{X}^{(i,k)})^{\mathrm{T}} \tag{4.20}$$

$$\mathbf{B}^{(i,k)} = \mathbf{V}^{(i,k)}\mathbf{S}^{(i,k)}(\mathbf{X}^{(i,k)})^{\mathrm{T}}, \tag{4.21}$$

where $\mathbf{U}^{(i,k)} \in \mathbb{R}^{n_2\times n_2}$, $\mathbf{V}^{(i,k)} \in \mathbb{R}^{n_2\times n_2}$ are unitary matrices, $\mathbf{X}^{(i,k)} \in \mathbb{R}^{n_2\times n_2}$ is non-singular, and $\mathbf{C}^{(i,k)} = [\mathbf{\Lambda}_{\mathbf{A}}^{(i,k)}\, \mathbf{0}]$ and $\mathbf{S}^{(i,k)} = [\mathbf{\Lambda}_{\mathbf{B}}^{(i,k)}\, \mathbf{0}]$ are diagonal matrices with

$$\mathbf{\Lambda}_{\mathbf{A}}^{(i,k)} = \begin{array}{c} \\ {\scriptstyle r_{\mathbf{x}_1}^{(i,k)}+r_{\mathbf{x}_2}^{(i,k)}-r_{\mathbf{x}}^{(i,k)}} \\ {\scriptstyle r_{\mathbf{x}}^{(i,k)}-r_{\mathbf{x}_1}^{(i,k)}} \end{array} \begin{array}{c} \begin{array}{cc} {\scriptstyle r_{\mathbf{x}_1}^{(i,k)}+r_{\mathbf{x}_2}^{(i,k)}-r_{\mathbf{x}}^{(i,k)}} & {\scriptstyle r_{\mathbf{x}}^{(i,k)}-r_{\mathbf{x}_1}^{(i,k)}} \end{array} \\ \begin{bmatrix} \mathbf{I} & \\ & \mathbf{D}_{\mathbf{A}}^{(i,k)} \end{bmatrix} \end{array}, \tag{4.22}$$

$$\mathbf{\Lambda}_{\mathbf{B}}^{(i,k)} = \begin{array}{c} \\ {\scriptstyle r_{\mathbf{x}_1}^{(i,k)}+r_{\mathbf{x}_2}^{(i,k)}-r_{\mathbf{x}}^{(i,k)}} \\ {\scriptstyle r_{\mathbf{x}_1}^{(i,k)}+r_{\mathbf{x}_2}^{(i,k)}-r_{\mathbf{x}}^{(i,k)}} \end{array} \begin{array}{c} \begin{array}{cc} {\scriptstyle r_{\mathbf{x}_1}^{(i,k)}+r_{\mathbf{x}_2}^{(i,k)}-r_{\mathbf{x}}^{(i,k)}} & {\scriptstyle r_{\mathbf{x}}^{(i,k)}-r_{\mathbf{x}_1}^{(i,k)}} \end{array} \\ \begin{bmatrix} \mathbf{0} & \\ & \mathbf{D}_{\mathbf{B}}^{(i,k)} \end{bmatrix} \end{array}, \tag{4.23}$$

where we have reported explicitly the dimensions of the square blocks in $\mathbf{\Lambda}_{\mathbf{A}}^{(i,k)}$ and $\mathbf{\Lambda}_{\mathbf{B}}^{(i,k)}$

We then define the matrix

$$\bar{\mathbf{\Phi}}_2 = [(\bar{\mathbf{\Phi}}_2^{(1,1)})^{\mathrm{T}}, \ldots, (\bar{\mathbf{\Phi}}_2^{(K_1,K_2)})^{\mathrm{T}}]^{\mathrm{T}} \tag{4.24}$$

and the matrices

$$\bar{\mathbf{\Phi}}_2^{(i,k)} = \left[\mathbf{I}_{r_{\mathbf{x}_1}^{(i,k)}+r_{\mathbf{x}_2}^{(i,k)}-r_{\mathbf{x}}^{(i,k)}}\ \mathbf{0}\right](\mathbf{X}^{(i,k)})^{-1}. \tag{4.25}$$

Finally, the suboptimal measurement matrix $\mathbf{\Phi}_2$ is given by picking any m_2 rows from $\bar{\mathbf{\Phi}}_2$. More specifically, $\mathbf{\Phi}_2$ is obtained by picking $m_2^{(i,k)}$ rows from $\bar{\mathbf{\Phi}}_2^{(i,k)}$ so that $\sum_{i,k} m_2^{(i,k)} = m_2$.[2]

In fact, the $m_2^{(i,k)}$ measurements picked from $\bar{\mathbf{\Phi}}_2^{(i,k)}$ capture the portion of the linear space space spanned by signals drawn from the distribution $p(\mathbf{x}_2|C_1 = i, C_2 = k)$ which is not occupied by signals drawn from the distri-

[2]It is also possible to tighten further the sufficient conditions in Theorem 28 by choosing the values of $m_2^{(i,k)}$ in order to minimize the value of the maximum among the right hand side of (4.17) for all (i,k), subject to the constraints $m_2^{(i,k)} \leq r_{\mathbf{x}_1}^{(i,k)} + r_{\mathbf{x}_2}^{(i,k)} - r_{\mathbf{x}}^{(i,k)}$, $\forall(i,k)$ and $\sum_{i,k} m_2^{(i,k)} = m_2$.

bution $p(\mathbf{x}_2|\mathbf{x}_1, C_1 = i, C_2 = k)$. The remaining $m_2 - m_2^{(i,k)}$ measurements act as random measurements for signals in class $C_1 = i, C_2 = k$.

Via Theorem 28 and 21 it is also possible to showcase the merit of an optimal measurement matrix $\mathbf{\Phi}_2^\star$ in comparison to a random one, in terms of the number of measurements that guarantee error free reconstruction in the low-noise regime.

Corollary 29. *Consider the measurement and source model in Theorem 28. Let $\mathcal{M}_{\mathsf{GMM}}^{\mathsf{R}}$ be the set of pairs (m_1, m_2) that verify the sufficient conditions for $\lim_{\sigma_1^2,\sigma_2^2\to 0} MMSE(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2) = 0$ in Theorem 21 for random $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, and let $\mathcal{M}_{\mathsf{GMM}}^{\mathsf{D}}$ be the set of pairs (m_1, m_2) that verify the sufficient conditions for $\lim_{\sigma_1^2,\sigma_2^2\to 0} MMSE(\sigma_1^2, \sigma_2^2, \mathbf{\Phi}_1, \mathbf{\Phi}_2^\star) = 0$ in Theorem 28, for random $\mathbf{\Phi}_1$ and optimal $\mathbf{\Phi}_2^\star$. Then, it holds*

$$\mathcal{M}_{\mathsf{GMM}}^{\mathsf{R}} \subseteq \mathcal{M}_{\mathsf{GMM}}^{\mathsf{D}}. \tag{4.26}$$

Proof: See Appendix B.4. ■

We finally note that, for both cases of Gaussian and GMM sources, designed projection kernels $\mathbf{\Phi}_2$ allow to capture the information contained in the side information $\mathbf{x}_2$ which is mostly correlated with $\mathbf{x}_1$. In fact, designed kernels aim at neglecting the information contained in the space spanned by signals drawn from the distribution $p(\mathbf{x}_2|\mathbf{x}_1)$, which can be interpreted as the innovation component of $\mathbf{x}_2$ with respect to $\mathbf{x}_1$.

4.4 Numerical Results

We now aim to show how numerical results with synthetic data align well with our theoretical derivations. We also use the real imaging data to showcase the impact of projection design using our proposed schemes. We first provide numerical results with synthetic data drawn from both Gaussian and GMM, that highlight the value of using designed measurement kernels with respect to random ones. We then consider similar trends for a reconstruction example with real-world imaging. In all experiments, we set $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

4.4.1 Synthetic Data: Gaussian Sources

We start by considering the simpler case where signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are described by a joint single Gaussian distribution with $K_1 = K_2 = 1$, and with dimensions $n_1 = 14$ and $n_2 = 6$, respectively. We also consider that the means associated with the Gaussian distribution of $\mathbf{x}_1$ and $\mathbf{x}_2$ are assumed to be zero, i.e. $\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{0}$, and the covariance matrices are independently drawn from a continuous pdf over the corresponding Grassmann manifold such that $r_{\mathbf{x}} = 5$, $r_{\mathbf{x}_1} = 4$ and $r_{\mathbf{x}_2} = 3$. We assume that the measurement matrix $\mathbf{\Phi}_1$ is obtained randomly with i.i.d., zero-mean, Gaussian entries with fixed variance. We also fix the number of measurements associated to the side information signal, $m_2 = 2$, and we consider the following scenarios: i) the case when $\mathbf{\Phi}_2$ is generated randomly with i.i.d., zero-mean, Gaussian entries with fixed variance; ii) the case when $\mathbf{\Phi}_2$ is generated numerically by addressing the optimization problem[3] in (4.8); iii) the case when $\mathbf{\Phi}_2$ is generated via the projection kernel design in (4.14).

Fig. 4.3 illustrates the MMSE vs. $1/\sigma^2$. We can note that the MMSE tends to zero when number of measurements $m_1 \geq 3$ and $m_2 = 2$ for the case of random projection kernels, thus matching the conditions Theorem 20. In contrast, we can observe the MMSE approaches zero in the low-noise regime when $m_1 \geq 2$ and $m_2 = 2$ for both designs in (4.8) and (4.14), which aligns very well with the predictions in Theorem 26. It is also obvious that the number of projections that guarantee the MMSE phase transition is less for scenarios where one uses designed measurements in comparison to scenarios where one uses random measurements. In particular, both kernel designs ii) and iii) provide similar values of MMSE, where the design in (4.14) requires the same number of projections as the optimal design in (4.8).

We now consider a joint Gaussian distribution with $K_1 = K_2 = 1$, for high-dimensional signals $\mathbf{x}_1$ and $\mathbf{x}_2$ with $n_1 = 100$ and $n_2 = 20$, respectively. We also consider that both signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are zero-mean, i.e., $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \mathbf{0}$, and the covariance matrices are independently drawn from a continuous pdf over the corresponding Grassmann manifold such that $r_{\mathbf{x}} = 5$, $r_{\mathbf{x}_1} = 4$ and $r_{\mathbf{x}_2} = 3$. In

[3] We consider the gradient descent optimization algorithm for finding the minimum error iteratively.

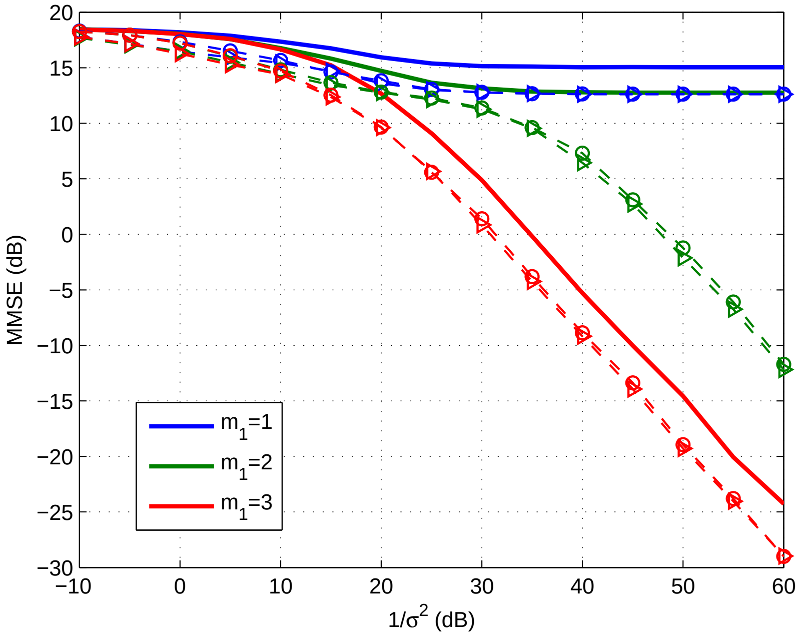


Figure 4.3: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ and $m_2 = 2$ for joint Gaussian sources. Side information with random projection kernel (solid lines). Side information with designed projection kernel $\mathbf{\Phi}_2^\star$ (numerical solution) (triangles) and suboptimal design (based on GSVD) (circles).

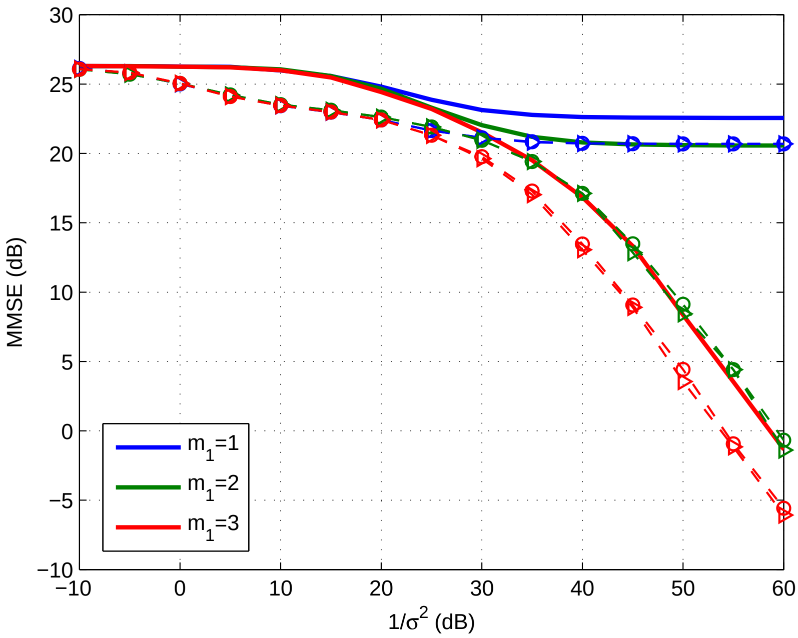


Figure 4.4: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3$ and $m_2 = 2$ for joint Gaussian sources for high-dimensional signals $\mathbf{x}_1$ and $\mathbf{x}_2$. Side information with random projection kernel (solid lines). Side information with designed projection kernel $\mathbf{\Phi}_2^\star$ (numerical solution) (triangles) and suboptimal design (based on GSVD) (circles).

this experiment, we also assume that the measurement matrix $\mathbf{\Phi}_1$ is generated randomly with i.i.d., zero-mean, Gaussian entries with fixed variance. We fix the number of measurements associated to the side information signal, $m_2 = 2$, and we consider three different scenarios: i) the case when $\mathbf{\Phi}_2$ is generated randomly with i.i.d., zero-mean, Gaussian entries with fixed variance; ii) the case when $\mathbf{\Phi}_2$ is generated numerically by addressing the optimization problem in (4.8); iii) the case when $\mathbf{\Phi}_2$ is generated via the projection kernel design in (4.14).

Fig. 4.4 demonstrates the MMSE vs. $1/\sigma^2$. We can note that the MMSE tends to zero in the low-noise regime when $m_1 \geq 3$ and $m_2 = 2$ for the case of random projection kernels, as predicted by Theorem 20. On the other hand, we observe that designed projection kernels guarantee that the reconstruction MMSE approaches zero in the low-noise regime when $m_1 \geq 2$ and $m_2 = 2$, thus matching the conditions proposed by Theorem 26. Fig. 4.4 also depicts that the kernel design associated with the side information leads to reliable reconstruction with a lower number of measurements with respect to the random kernels for high-dimensional signals. Moreover, we also observe that, in this case, the design in (4.14) requires the same number of projections as the optimal design in (4.8).

4.4.2 Synthetic Data: GMM Sources

We now consider the more general case in which signals are drawn from a joint GMM distribution. Here, we consider both signals $\mathbf{x}_1$ and $\mathbf{x}_2$, with dimensions $n_1 = 14$ and $n_2 = 6$, and $K_1 = K_2 = 2$. All means associated with the various Gaussian distributions are zero, i.e., $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \mathbf{0}$, and the covariance matrices are randomly generated such that $r_{\mathbf{x}}^{(i,k)} = 5$, $r_{\mathbf{x}_1}^{(i,k)} = 3$ and $r_{\mathbf{x}_2}^{(i,k)} = 3$ for $i = 1, 2$ and $k = 1, 2$. The images of the covariance matrices associated with different classes are drawn uniformly at random from the corresponding Grassmann manifold. We also assume that the projection kernel $\mathbf{\Phi}_1$ has i.i.d., zero-mean, Gaussian entries with fixed variance. We fix the number of side information measurements $m_2 = 2$ and we also consider three different cases: i) the case when $\mathbf{\Phi}_2$ is random, with i.i.d., zero-mean, Gaussian entries with fixed variance; ii) the case when $\mathbf{\Phi}_2$ is obtained via numerical solution of the problem (4.8); iii) the case when $\mathbf{\Phi}_2$ is obtained via the projection kernel design in (4.25).

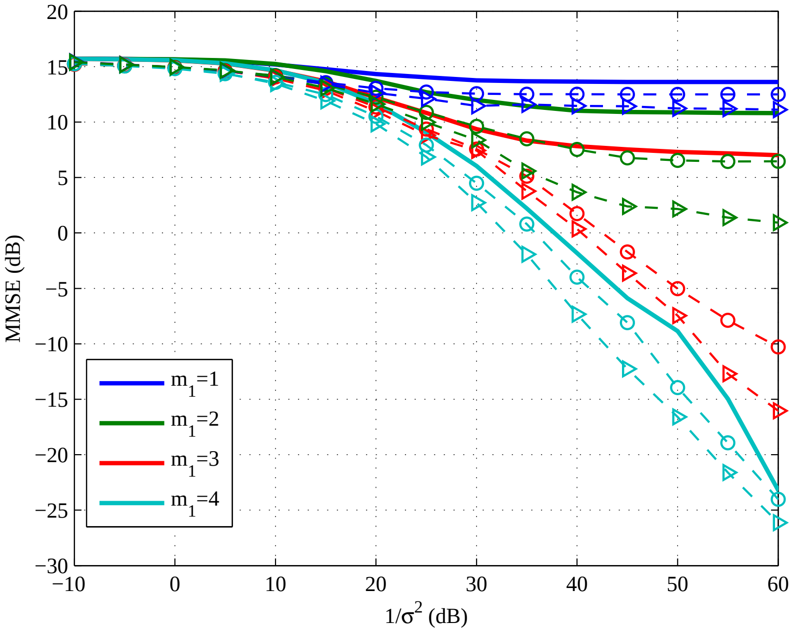


Figure 4.5: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3, 4$ and $m_2 = 2$ for joint GMM sources. Side information with random projection kernel (solid lines). Side information with designed projection kernel $\mathbf{\Phi}_2^\star$ (numerical solution) (triangles) and suboptimal design (based on GSVD) (circles).

Fig. 4.5 reports the MMSE vs. $1/\sigma^2$. We observe that random projection kernels guarantee reliable reconstruction when $m_1 + m_2 > r_{\mathbf{x}}^{(i,k)} = 5$, as predicted by the results in Theorem 21[4]. On the other hand, designing the projection kernel $\mathbf{\Phi}_2$ achieves reliable reconstruction with a lower number of measurements, as the MMSE is observed to approach zero in the low-noise regime when $m_1 > 2$ and $m_2 = 2$. Such behavior is well aligned with the sufficient condition expressed in Theorem 28, since $r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)} = 2$, $\forall (i,k)$. We can also note that both designs in (4.25) and (4.8) reduce significantly the minimum number of measurements required for the MMSE to approach zero in the low-noise regime in relation to the case of random kernels. Moreover, we observe that, in this case, it turns out that the design in (4.25) requires the same number of projections as the optimal design in (4.8) in order to guarantee reliable reconstruction, but this may not always be the case.

[4]The analysis carried out in [81] considers a slightly different signal model in which $\mathbf{x}_1$ and $\mathbf{x}_2$ obey only an approximately low-rank joint GMM distribution and linear measurements are noiseless. However, the results on the number of random measurements required for reliable reconstruction can be easily modified to fit the signal model considered in this work.

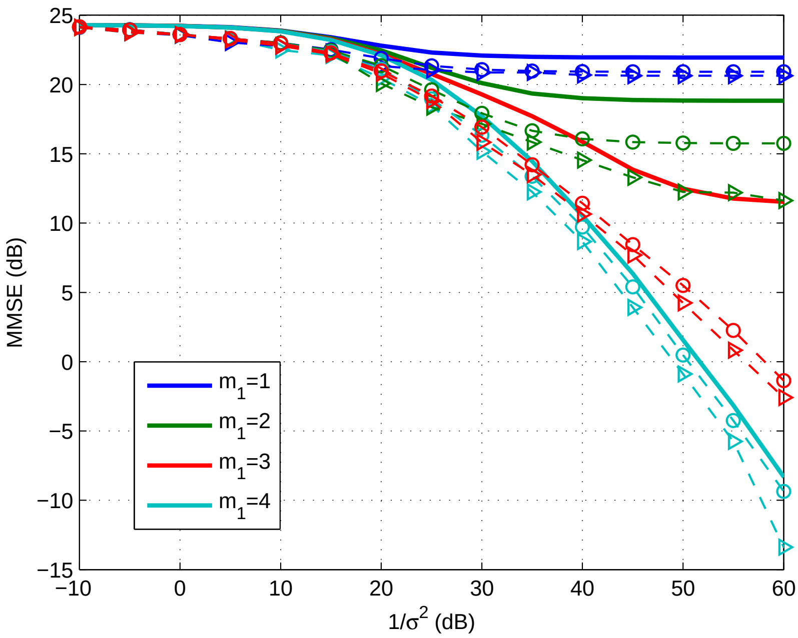


Figure 4.6: MMSE vs. $1/\sigma^2$ for $m_1 = 1, 2, 3, 4$ and $m_2 = 2$ for joint GMM sources. Side information with random projection kernel (solid lines). Side information with designed projection kernel $\mathbf{\Phi}_2^\star$ (numerical solution) (triangles) and suboptimal design (based on GSVD) (circles).

We then consider a joint GMM distribution for signals $\mathbf{x}_1$ and $\mathbf{x}_2$ with higher dimensions $n_1 = 100$ and $n_2 = 20$, and $K_1 = K_2 = 2$. We also consider that both signals $\mathbf{x}_1$ and $\mathbf{x}_2$ are zero-mean, i.e., $\boldsymbol{\mu}_{\mathbf{x}}^{(i,k)} = \mathbf{0}$, and the covariance matrices are randomly generated such that $r_{\mathbf{x}}^{(i,k)} = 5$, $r_{\mathbf{x}_1}^{(i,k)} = 3$ and $r_{\mathbf{x}_2}^{(i,k)} = 3$ for $i = 1, 2$ and $k = 1, 2$. The images of the covariance matrices associated with different classes are drawn uniformly at random from the corresponding Grassmann manifold. We also assume that the projection kernel $\mathbf{\Phi}_1$ has i.i.d., zero-mean, Gaussian entries with fixed variance. We fix the number of side information measurements $m_2 = 2$ and we also consider three different cases: i) the case when $\mathbf{\Phi}_2$ is random, with i.i.d., zero-mean, Gaussian entries with fixed variance; ii) the case when $\mathbf{\Phi}_2$ is obtained via numerical solution of the problem (4.8); iii) the case when $\mathbf{\Phi}_2$ is obtained via the projection kernel design in (4.25).

Fig. 4.6 reports the reconstruction MMSE vs. $1/\sigma^2$. We can observe that the MMSE tends to zero in the low-noise regime when number of measurements $m_1 + m_2 > r_{\mathbf{x}}^{(i,k)} = 5$ for the case of random projection kernels (solid lines). Such behavior is well aligned with the condition expressed in Theorem 21. On the other hand, the design of projection kernel associated with the side information guarantees the reliable reconstruction when $m_1 > 2$ and $m_2 = 2$, as predicted by Theorem 28. We can also note that both designs in (4.8) and (4.25) can reduce significantly the minimum number of measurements needed for the MMSE phase transition with respect to the case of random kernels. Moreover, the design in (4.8) requires the same number of measurements as the optimal design in (4.25).

4.4.3 Real Data

Finally, we use the real-world images to showcase whether the measurement designs identified in (4.8) and (4.25) can lead to better reconstruction performance. We consider a high-resolution image Lena with dimension 512×512 as the signal of interest, and a low-resolution image (128×128 pixels) of the same subject as side information, as depicted in Fig. 4.7. In particular, they are partitioned into non-overlapping patches, so that we consider vectors $\mathbf{x}_1$ correspond to 8×8 non-overlapping patches from the signal of interest, while vectors $\mathbf{x}_2$ correspond to 2×2 non-overlapping patches from the lower resolution signal, that are divided into overall 4096 patches for both images. We also assume that vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint GMM distribution, with $K_1 = K_2 = 20$. All parameters associated to each Gaussian distribution within the joint GMM, include prior class probabilities, class-conditioned means and class-conditioned covariance matrices, are trained via the EM algorithm over a set of patches extracted from images in the Caltech 101 dataset [97]. We assume the projection matrix $\mathbf{\Phi}_1$ is obtained randomly with i.i.d., zero-mean, Gaussian entries with fixed variance.

The results in Fig. 4.8 report the reconstruction images and highlight the corresponding PNSR values, with the fixed number of linear measurements $m_1 = 10$ and $m_2 = 4$, and with noise level $\sigma^2 = -40$ dB. From the top to the bottom, reconstruction results are shown with different projection matrices associated to the side information: (a) when $\mathbf{\Phi}_2$ is random with i.i.d., zero-mean, Gaussian entries with fixed variance; (b) when $\mathbf{\Phi}_2$ is obtained according to the design in (4.8); and (c) when $\mathbf{\Phi}_2$ is obtained by the solution of optimal problem in (4.25)[5].

[5]In the last case, the matrix $\bar{\mathbf{\Phi}}_2 \in \mathbb{R}^{80\times 4}$ was generated by picking the first 4 rows of the matrix $(\mathbf{X}^{(i,k)})^{-1}$ in (4.25), for all (i,k). Then, $\mathbf{\Phi}_2$ was obtained by picking the $m_2 = 4$ rows from $\bar{\mathbf{\Phi}}_2$ that corresponded to the lowest reconstruction MMSE. In general, $\mathbf{\Phi}_2$ could be formed by randomly picking $m_2 = 4$ rows from $\bar{\mathbf{\Phi}}_2$.

Figure 4.7: Left: input signal, high-resolution image (512×512). Right: side information, low-resolution image (128×128)

We can observe that the reconstruction results for both design approaches in Fig. 4.8 (b) and (c) can achieve approximately 2.5 dB gains of PSNR with respect to the case of random projection kernels as shown in Fig. 4.8 (a). Moreover, we also observe that the design described in (4.25) achieves similar PSNR values as when the projection kernel is designed via more computationally expensive numerical solution of the optimization problem in (4.8). This suggests that (4.25) can be used as an off-the-shelf measurement design procedure applicable to compressive sensing systems with side information where the signals are modeled via GMMs.

We then study the case when vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are drawn from a joint GMM distribution, with $K_1 = K_2 = 10$ and $K_1 = K_2 = 30$. As in previous experiment, all parameters associated to each Gaussian distribution within the joint GMM, containing prior class probabilities, class-conditioned means and class-conditioned covariance matrices, are trained by using the EM algorithm over a set of patches extracted from images in the "Caltech 101" dataset. In this experiment, we consider both images are then partitioned into non-overlapping patches, so that the input vectors $\mathbf{x}_1$ correspond to 8×8 non-overlapping patches, while the vectors $\mathbf{x}_2$ correspond to with non-overlapping 2×2, that are partitioned into overall 4096 patches for both images. Here, we also assume projection matrix $\mathbf{\Phi}_1$ is random with i.i.d., zero-mean, Gaussian entries with fixed variance.

Figs. 4.9 and 4.10 represent the reconstruction images with corresponding PSNR values with the fixed number of linear measurements $m_1 = 10$ and $m_2 = 4$, and with noise level $\sigma^2 = -40$, for $K_1 = K_2 = 10$ and $K_1 = K_2 = 30$ respectively. From the top to the bottom, both reconstruction results are presented with different measurements designs: (a) $\mathbf{\Phi}_2$ is obtained randomly with i.i.d., zero-mean, Gaussian entries with fixed variance; (b) $\mathbf{\Phi}_2$ is generated via the design in (4.8); and (c) $\mathbf{\Phi}_2$ is generated using (4.25).

The results in Figs. 4.9(b) and (c), show both designs provide similar values of PSNR, exhibiting approximately 2.5 dB gains, with respect to the case of random projection kernels as shown in Fig. 4.9(a). Figs. 4.10(b) and (c) lead to a PSNR gain of approximately 2.5 dB in comparison to case of random projection kernels as shown in Fig. 4.10(a). As in previous experiments, the designs in (4.8) and (4.25) can lead to better reconstruction performance with respect to random projection kernels. In this experiment, the result also suggests that the richer GMM can lead to greater reconstruction performance.

Moreover, we also provide various reconstruction PSNR results with different number of noise levels with different number of measurements $m_1 = 5, 10, 15$ (and we fix $m_2 = 4$) and different number of classes, as reported in Tables 4.1, 4.2, and 4.3. We can observe that the reconstruction PSNR are improved with increasingly number of measurements and dropping noise levels. Furthermore, designing projection kernel in (4.8) and (4.25) can lead to significant reconstruction PSNR gains in relation to the random measurement kernels. The results also suggest that the patches described by a 30-classes joint GMM distribution yields PSNR values slightly higher than those obtained with 20-classes and 10-classes joint GMM distribution.

Finally, we consider the case when the signal of interest and the side information are divided into larger size patches, where the vectors $\mathbf{x}_1$ correspond to 16×16 non-overlapping patches, while the vectors $\mathbf{x}_2$ correspond to non-overlapping 4×4 extracted from Fig. 4.7 respectively, so the overall number of patches is 1024 for both images. In this experiment, we consider the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are jointly described by a GMM distribution with $K_1 = K_2 = 20$, and all parameters associated to each Gaussian distribution are obtained via the EM algorithm using

Table 4.1: PSNR values (dB) of the reconstruction with $K_1 = K_2 = 10$ with number of measurements $m_1 = 5, 10, 15$.

SNR	−20 dB	−30 dB	−40 dB	−50 dB	−60 dB
			$m_1 = 5$		
case (a)	17.7547	19.8029	22.9689	24.4912	25.7977
case (b)	25.7588	25.7842	26.9184	27.8746	28.0723
case (c)	25.6211	25.7652	26.9172	27.8650	28.0725
			$m_1 = 10$		
case (a)	18.8661	22.6112	25.3438	26.9296	27.6130
case (b)	25.8054	26.1864	27.9296	28.2594	29.3282
case (c)	25.8015	26.1541	27.9288	28.2545	29.3253
			$m_1 = 15$		
case(a)	21.6150	25.8422	27.5346	28.3414	30.4292
case(b)	27.7636	27.7771	28.1200	28.7851	30.4387
case(c)	27.6184	27.7631	28.1197	28.7827	30.4317

Table 4.2: PSNR values (dB) of the reconstruction with $K_1 = K_2 = 20$ with number of measurements $m_1 = 5, 10, 15$.

SNR	−20 dB	−30 dB	−40 dB	−50 dB	−60 dB
			$m_1 = 5$		
case (a)	17.9117	20.1615	23.3938	26.4714	27.4714
case (b)	26.0784	26.8008	27.6544	28.3922	29.3227
case (c)	26.0070	26.8961	27.6534	28.3917	29.3218
			$m_1 = 10$		
case (a)	19.1011	22.8247	25.4820	27.8020	29.1471
case (b)	27.5887	27.6190	28.1036	28.9435	30.4298
case (c)	27.5849	27.6117	28.1028	28.9302	30.4321
			$m_1 = 15$		
case (a)	21.6980	25.8760	27.6360	28.4098	30.7600
case (b)	27.8837	27.8991	28.1466	28.9542	30.7652
case (c)	27.8102	27.8951	28.1448	28.9530	30.7636

images from the "Caltech 101" dataset [97]. We also assume the projection kernel $\mathbf{\Phi}_1$ is generated with i.i.d., zero-mean, Gaussian entries with fixed variance.

Fig. 4.11 reveals the reconstruction images with corresponding PSNR values, with the fixed number of linear measurements $m_1 = 10$ and $m_2 = 4$, and with noise level $\sigma^2 = -40$ dB. From the top to the bottom, reconstructions are presented with different measurements designs: (a) $\mathbf{\Phi}_2$ is obtained randomly with i.i.d., zero-mean, Gaussian entries with fixed variance; (b) $\mathbf{\Phi}_2$ is generated numerically via the solution of (4.8); and (c) $\mathbf{\Phi}_2$ is generated using (4.25).

Table 4.3: PSNR values (dB) of the reconstruction with $K_1 = K_2 = 30$ with number of measurements $m_1 = 5, 10, 15$.

SNR	−20 dB	−30 dB	−40 dB	−50 dB	−60 dB
	$m_1 = 5$				
case(a)	17.9421	23.1869	26.7255	28.5371	29.4808
case(b)	26.1324	27.2742	27.9385	28.7579	29.5099
case(c)	26.0545	27.2691	27.8026	28.7545	29.5073
	$m_1 = 10$				
case(a)	22.0896	22.8443	25.5192	28.2573	30.5780
case(b)	27.6039	27.8175	28.1402	29.0805	30.6794
case(c)	27.5548	27.8068	28.1392	29.0742	30.6768
	$m_1 = 15$				
case(a)	22.6783	25.9568	27.7262	29.0883	31.0636
case(b)	27.9559	27.9687	28.6440	29.3098	31.0382
case(c)	27.8656	27.9629	28.2406	29.1044	31.0310

Both designs in Figs. 4.11 (b) and (c) provide PSNR gains of approximately 2 dB with respect to the case of random projections as shown in Figs. 4.8 (a). In particular, we also observe the design in (4.25) and the design via the approximating numerically the solution of (4.8) provide similar reconstruction performance. In this experiment, however, patches extracted with larger size from both the signal of interest and the side information does not enhance the reconstruction performance with respect to our previous experiment. This is due to the fact that the GMM is not rich enough to describe such larger dimensional patches well.

For a higher dimensional patches, we also consider the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are jointly described by a GMM distribution with $K_1 = K_2 = 30$. Fig. 4.12 reveals the reconstruction images with corresponding PSNR values, with the fixed number of linear measurements $m_1 = 10$ and $m_2 = 4$, and with noise level $\sigma^2 = -40$ dB. The results in Fig. 4.12 also show that both designs can improve the reconstruction quality with respect to random projection kernels. We can also observe that the patches described by a 30-classes joint GMM distribution yields PSNR values slightly higher than those obtained with 20-classes joint GMM distribution.

(a) Random kernels.

(b) Designed kernel (numerical solutions).

(c) Designed kernel (GSVD).

Figure 4.8: Reconstruction results of image "Lena" for $\sigma^2 = -40$ dB with $K_1 = K_2 = 20$. From the top to the bottom, the reconstruction PSNR values are 25.4820 dB, 28.1036 dB, and 28.1028 dB.

(a) Random kernels.

(b) Designed kernel (numerical solutions).

(c) Designed kernel (GSVD).

Figure 4.9: Reconstruction results of image "Lena" for $\sigma^2 = -40$ dB with $K_1 = K_2 = 10$. From the top to the bottom, the reconstruction PSNR values are 25.3438 dB, 27.9296 dB, and 27.9288 dB.

(a) Random kernels.

(b) Designed kernel (numerical solutions).

(c) Designed kernel (GSVD).

Figure 4.10: Reconstruction results of image "Lena" for $\sigma^2 = -40$ dB with $K_1 = K_2 = 30$. From the top to the bottom, the reconstruction PSNR values are 25.5192 dB, 28.1402 dB, and 28.1392 dB.

(a) Random kernels.

(b) Designed kernel (numerical solutions).

(c) Designed kernel (GSVD).

Figure 4.11: Reconstruction results of image "Lena" for $\sigma^2 = -40$ dB with $K_1 = K_2 = 20$. From the top to the bottom, the reconstruction PSNR values are 23.6680 dB, 25.7870 dB, and 25.3023 dB.

(a) Random kernels.

(b) Designed kernel (numerical solutions).

(c) Designed kernel (GSVD).

Figure 4.12: Reconstruction results of image "Lena" for $\sigma^2 = -40$ dB with $K_1 = K_2 = 30$. From the top to the bottom, the reconstruction PSNR values are 23.9159 dB, 26.2042 dB, and 26.0361 dB.

4.5 Summary

In this work, we have assessed the impact of kernel design on reconstruction performance in the presence of side information, by assuming that both the signal of interest and side information are jointly drawn from both Gaussian and GMM distributions, with possibly low-rank, class-conditioned covariance matrices. In particular, we have considered the case when the projection kernel associated to the signal of interest is random, whereas the projection kernel associated to the side information is designed. The case for random projection kernels is also used for comparison purposes.

We have provided sufficient (and necessary in some cases) conditions on the number of linear projections for reliable reconstruction in the low-noise regime. Our proposed theorems have shown that these proposed designs capture the information contained in the side information which is mostly correlated to the signal of interest. The results have shown that our proposed designs can lead to significant improvement of reconstruction performance for both Gaussian and GMM sources.

We have also provided numerical experiments with synthetic data for both Gaussian and GMM signals to confirm our proposed theorems align well with proposed theorems. The results with both Gaussian and GMM sources have illustrated that can lead to significant reduction in terms of reconstruction MMSE for finite noise levels. In particular, careful design of $\mathbf{\Phi_2}$ can lead to a reduction of the minimum number of measurements needed for reliable reconstruction in the low-noise regime. Finally, we also show that our proposed designs have the potential to significantly improve the performance of a real-world imaging, exhibiting PSNR gains of approximately 2.5 dB with respect to the case of random projection kernels.

Appendices

Appendix B

Proofs and Derivations of Chapters 4

B.1 Proof of Theorem 26

We now provide necessary and sufficient conditions on the number of measurements m_1 and m_2 for the MMSE to approach zero in low noise regime.

We will use the symbol $\text{MMSE}(\mathbf{x}|\mathbf{y})$ to explicitly denote the MMSE in recovering $\mathbf{x}$ from the observation of $\mathbf{y}$. Then, by noting that the reconstruction MMSE does not depend on the value of the mean of the input signal, and by taking the expectation in the MMSE expression with respect to the random variables $\mathbf{x}_1|\mathbf{y}_2$ and $\mathbf{y}_2$, separately, it is possible to show that

$$\text{MMSE}(\mathbf{x}_1|\mathbf{y}_1, \mathbf{y}_2) = \text{MMSE}(\mathbf{z}|\mathbf{\Phi}_1\mathbf{z} + \mathbf{w}_1), \tag{B.1}$$

where $\mathbf{z} \sim p(\mathbf{x}_1|\mathbf{y}_2)$. By leveraging [93], we observe that $\text{MMSE}(\mathbf{z}|\mathbf{\Phi_1}\mathbf{z} + \mathbf{w}_1)$ tends to be zero as $\sigma_1^2 \to 0$, if and only if

$$m_1 \geq \text{rank}(\mathbf{\Sigma_z}) = r_{\mathbf{z}}, \tag{B.2}$$

Note that the covariance matrix $\mathbf{\Sigma_z}$ can be written as

$$\mathbf{\Sigma_z} = \mathbf{\Sigma}_{\mathbf{x}_1} - \mathbf{\Sigma}_{\mathbf{x}_{12}}\mathbf{\Phi}_2^{\mathrm{T}}(\mathbf{\Phi}_2\mathbf{\Sigma}_{\mathbf{x}_2}\mathbf{\Phi}_2^{\mathrm{T}} + \mathbf{I}\sigma_2^2)^{\dagger}\mathbf{\Phi}_2\mathbf{\Sigma}_{\mathbf{x}_{21}}, \tag{B.3}$$

and in particular, when $\sigma_2^2 \to 0$, $\mathbf{\Sigma_z}$ tends to

$$\mathbf{\Sigma_z} = \mathbf{\Sigma}_{\mathbf{x}_1} - \mathbf{\Sigma}_{\mathbf{x}_{12}} \mathbf{\Phi}_2^{\mathrm{T}} (\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}})^{\dagger} \mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_{21}}. \tag{B.4}$$

Note that $\mathbf{\Sigma_z}$ in (B.4) is the generalized Schur complement [99] of the block $\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}}$ of the positive semidefinite matrix

$$\mathbf{\Sigma}_{\mathbf{x}_1 \mathbf{\Phi}_2 \mathbf{x}_2} = \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{x}_1} & \mathbf{\Sigma}_{\mathbf{x}_{12}} \mathbf{\Phi}_2^{\mathrm{T}} \\ \mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_{21}} & \mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}} \end{bmatrix}, \tag{B.5}$$

so that we have [99]

$$\mathrm{rank}(\mathbf{\Sigma}_{\mathbf{x}_1 \mathbf{\Phi}_2 \mathbf{x}_2}) = r_{\mathbf{z}} + \mathrm{rank}(\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}}). \tag{B.6}$$

In addition, on considering the matrix

$$\mathbf{\Sigma}_{\mathbf{\Phi}_2 \mathbf{x}_2 \mathbf{x}_1} = \begin{bmatrix} \mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}} & \mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_{21}} \\ \mathbf{\Sigma}_{\mathbf{x}_{12}} \mathbf{\Phi}_2^{\mathrm{T}} & \mathbf{\Sigma}_{\mathbf{x}_1} \end{bmatrix}, \tag{B.7}$$

and on applying the same rank computation, we also have

$$\mathrm{rank}(\mathbf{\Sigma}_{\mathbf{\Phi}_2 \mathbf{x}_2 \mathbf{x}_1}) = r_{\mathbf{x}_1} + \mathrm{rank}\left(\mathbf{\Phi}_2 (\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}) \mathbf{\Phi}_2^{\mathrm{T}}\right). \tag{B.8}$$

By substituting (B.6) and (B.8), we can rewrite express $r_{\mathbf{z}}$ as

$$r_{\mathbf{z}} = r_{\mathbf{x}_1} - \mathrm{rank}(\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}}) + \mathrm{rank}(\mathbf{\Phi}_2 (\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}) \mathbf{\Phi}_2^{\mathrm{T}}). \tag{B.9}$$

The sufficient conditions for the MMSE phase transition can be obtained by considering the MMSE associated to a suboptimal projection matrix $\mathbf{\Phi}_2$ with respect to the optimization problem (4.8). Based on the GSVD of two matrices $\mathbf{\Sigma}_{\mathbf{x}_2}$ and $\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}$, the projection matrix $\mathbf{\Phi}_2$ is obtained in (4.14).

Then, it is straightforward to verify that $\mathrm{rank}(\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}}) = m_2'$ and $\mathrm{rank}(\mathbf{\Phi}_2 (\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}) \mathbf{\Phi}_2^{\mathrm{T}}) = 0$, thus implying that, when adopting the side information projection matrix $\mathbf{\Phi}_2$ in (4.14), the corresponding MMSE approaches zero in the low-noise regime if

$$m_1 \geq r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}, \tag{B.10}$$

which is equivalent to the condition (4.10).

Conversely, we can show that the condition in (B.10) is also necessary to achieve the MMSE phase transition when $\mathbf{\Phi}_2 = \mathbf{\Phi}_2^\star$. In order to do that, we show that, for any choice of $\mathbf{\Phi}_2 \in \mathbb{R}^{m_2 \times n_2}$, it holds

$$r_{\mathbf{z}} \geq r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}. \tag{B.11}$$

In particular, on leveraging the Sylvester's rank Theorem [101], which states that

$$\mathrm{rank}(\mathbf{AB}) = \mathrm{rank}(\mathbf{B}) - \dim(\mathrm{Im}(\mathbf{B}^{\mathrm{T}}) \cap \mathrm{Null}(\mathbf{A})), \tag{B.12}$$

we can write

$$\Delta(\mathbf{\Phi}_2) = \mathrm{rank}(\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2} \mathbf{\Phi}_2^{\mathrm{T}}) - \mathrm{rank}(\mathbf{\Phi}_2(\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}) \mathbf{\Phi}_2^{\mathrm{T}}) \tag{B.13}$$

$$= \dim(\mathrm{Im}(\mathbf{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}})) - \dim(\mathrm{Im}(\mathbf{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2})). \tag{B.14}$$

Then, on observing that $\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}$ is the generalized Schur complement of $\mathbf{\Sigma}_{\mathbf{x}_1}$ in $\mathbf{\Sigma}_{\mathbf{x}}$[102], and by leveraging [102, Lemma 4.1] in conjunction with [103, Theorem 4.3], we have that

$$\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}) \subseteq \mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}}), \tag{B.15}$$

and

$$\dim(\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2} - \mathbf{\Sigma}_{\mathbf{x}_{21}} \mathbf{\Sigma}_{\mathbf{x}_1}^{\dagger} \mathbf{\Sigma}_{\mathbf{x}_{12}})) = n_2 - r_{\mathbf{x}} + r_{\mathbf{x}_1} \tag{B.16}$$

$$\dim(\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2})) = n_2 - r_{\mathbf{x}_2}. \tag{B.17}$$

Finally, on leveraging the following lemma, we can prove that $\Delta(\mathbf{\Phi}_2) \leq \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}$.

Lemma 30. *Consider two linear spaces $\mathcal{D}_1, \mathcal{D}_2 \in \mathbb{R}^n$, such that $\mathcal{D}_2 \subseteq \mathcal{D}_1$, with dimensions d_1 and d_2, $d_1 \geq d_2$, respectively. Consider a third linear space $\mathcal{V} \in \mathbb{R}^n$, with dimension v. Then,*

$$\dim(\mathcal{V} \cap \mathcal{D}_1) - \dim(\mathcal{V} \cap \mathcal{D}_2) \leq \min\{v, d_1 - d_2\}. \tag{B.18}$$

Proof: It is straightforward to note that the difference in (B.18) is always upper bounded by v, since $\dim(\mathcal{V} \cap \mathcal{D}_1) \leq \dim(\mathcal{V}) = v$.

Then, we divide the proof into two cases. In the first case, we assume $\dim(\mathcal{V} \cap \mathcal{D}_1) \leq d_1 - d_2$, from which (B.18) follows immediately. In the second case, we assume $\dim(\mathcal{V} \cap \mathcal{D}_1) > d_1 - d_2$, and we write $\dim(\mathcal{V} \cap \mathcal{D}_1) = d_1 - d_2 + k$, with $k \in \{0, 1, \ldots, d_2\}$. Then, we have that

$$\begin{aligned} \dim(\mathcal{V} \cap \mathcal{D}_2) &\geq \dim((\mathcal{V} \cap \mathcal{D}_1) \cap \mathcal{D}_2) && \text{(B.19)} \\ &\geq [d_1 - d_2 + k + d_2 - d_1]^+ = k, && \text{(B.20)} \end{aligned}$$

where, in the last inequality, we have used fact that both $\mathcal{V} \cap \mathcal{D}_1$ and $\mathcal{D}_2$ are linear spaces contained in $\mathcal{D}_1$, which has dimension d_1. Therefore, from (B.20) we obtain $\dim(\mathcal{V} \cap \mathcal{D}_1) - \dim(\mathcal{V} \cap \mathcal{D}_2) \leq d_1 - d_2$, which concludes the proof of (B.18). ■

B.2 Proof of Corollary 27

Note that, given any value of m_2, the minimum value of m_1 that verifies the necessary and sufficient conditions for the MMSE to approach zero in the low-noise regime when both $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are random can be written as [81, Theorem 3]

$$m_1^{\mathsf{R}} = r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_2}\} + \min\{m_2, r_{\mathbf{x}} - r_{\mathbf{x}_1}\}, \tag{B.21}$$

whereas the minimum value of m_1 that verifies the necessary and sufficient condition in Theorem 22 can be written as

$$m_1^{\mathsf{D}} = r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\}. \tag{B.22}$$

Then, the proof of Corollary 27 is obtained by comparing (B.21) and (B.22) for different values of m_2:

- If $m_2 \leq r_{\mathbf{x}} - r_{\mathbf{x}_1}$, we have $m_1^{\mathsf{R}} = r_{\mathbf{x}_1} \geq r_{\mathbf{x}_1} - \min\{m_2, r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}\} = m_1^{\mathsf{D}}$;

- If $r_{\mathbf{x}} - r_{\mathbf{x}_1} < m_2 < r_{\mathbf{x}_2}$, we need to consider two further cases: if $m_2 \geq r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}$, then $m_1^{\mathsf{R}} = r_{\mathbf{x}} - m_2 > r_{\mathbf{x}} - r_{\mathbf{x}_2} = m_1^{\mathsf{D}}$; on the other hand, if $m_2 < r_{\mathbf{x}_1} + r_{\mathbf{x}_2} - r_{\mathbf{x}}$, then $m_1^{\mathsf{R}} = r_{\mathbf{x}} - m_2 \geq r_{\mathbf{x}_1} - m_2 = m_1^{\mathsf{D}}$.

- If $m_2 \geq r_{\mathbf{x}_2}$, then $m_1^{\mathsf{R}} = r_{\mathbf{x}} - r_{\mathbf{x}_2} = m_1^{\mathsf{D}}$.

Fig. B.1 illustrates with different shading colors the different sets of values (m_1, m_2) considered in the proof of Corollary 27.

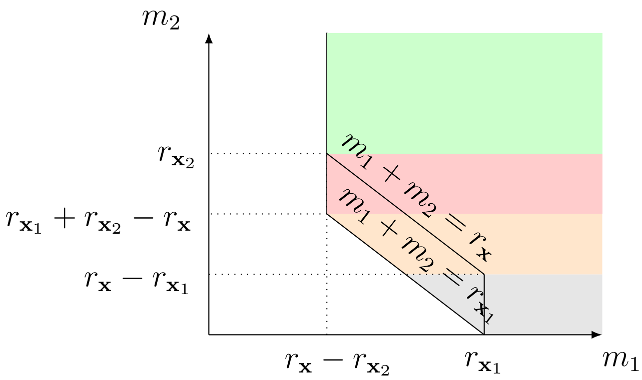


Figure B.1: Representation of the conditions on m_1 and m_2 for MMSE phase transition for Gaussian sources.

B.3 Proof of Theorem 28

When the signals $\mathbf{x}_1, \mathbf{x}_2$ are described via a joint GMM, the corresponding MMSE can be expressed as

$$\mathrm{MMSE}(\mathbf{x}_1|\mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}\left[\mathrm{MMSE}(\mathbf{z}|\mathbf{\Phi}_1\mathbf{z} + \mathbf{w}_1)\right], \tag{B.23}$$

where $\mathbf{z} \sim p(\mathbf{x}_1|\mathbf{y}_2)$ and where the outer expectation is taken with respect to $\mathbf{y}_2$. Note that, for a given value of $\mathbf{y}_2$, $p(\mathbf{z})$ is a GMM distribution with the same number of classes as the joint GMM $p(\mathbf{x}_1, \mathbf{x}_2)$, since it holds

$$p(\mathbf{z}) = \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} p(C_1 = i, C_2 = k | \mathbf{y}_2) \cdot p(\mathbf{x}_1 | \mathbf{y}_2, C_1 = i, C_2 = k) \tag{B.24}$$

$$= \sum_{i=1}^{K_1} \sum_{k=1}^{K_2} p(C_1 = i, C_2 = k | \mathbf{y}_2) \cdot \frac{p(\mathbf{x}_1, \mathbf{y}_2, C_1 = i, C_2 = k)}{p(\mathbf{y}_2, C_1 = i, C_2 = k)}. \tag{B.25}$$

Therefore, by leveraging [93, Theorem 3], we can conclude that a sufficient condition on the number of projections m_1 required for $\mathrm{MMSE}(\mathbf{z}|\mathbf{\Phi}_1\mathbf{z} + \mathbf{w}_1) \to 0$ when $\sigma_1^2 \to 0$ is given by

$$m_1 > \max_{(i,k)} \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)}) = \max_{(i,k)} r_{\mathbf{z}}^{(i,k)}, \tag{B.26}$$

where $\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)}$ is the covariance matrix of $\mathbf{z}$ conditioned on classes $C_1 = i, C_2 = k$. We also note that the covariances $\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)}$ depend only on the statistical description of $\mathbf{x}_1, \mathbf{y}_2$, and not on the particular realization of $\mathbf{y}_2$. Therefore, we can conclude that (B.26) represents also a sufficient condition for $\mathrm{MMSE}(\mathbf{x}_1|\mathbf{y}_1, \mathbf{y}_2) \to 0$ when $\sigma_1^2 \to 0$. The covariance matrix $\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)}$ can be written as

$$\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)} = \mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \mathbf{\Phi}_2^{\mathrm{T}} (\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} \mathbf{\Phi}_2^{\mathrm{T}} + \mathbf{I}\sigma_2^2)^{-1} \mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}. \tag{B.27}$$

Moreover, we are interested in determining the behavior of the MMSE in the low-noise regime, that is when both $\sigma_1^2 \to 0$ and $\sigma_2^2 \to 0$. In particular, when $\sigma_2^2 \to 0$, the matrices $\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)}$ can be expressed as

$$\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)} = \mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)} \mathbf{\Phi}_2^{\mathrm{T}} (\mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} \mathbf{\Phi}_2^{\mathrm{T}})^{\dagger} \mathbf{\Phi}_2 \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}, \tag{B.28}$$

due to continuity of the Schur component of a matrix with respect to its eigenvalues. Therefore, sufficient conditions for the MMSE to approach zero in the low-noise regime are obtained by expressing the quantities $r_{\mathbf{z}}^{(i,k)} = \mathrm{rank}(\mathbf{\Sigma}_{\mathbf{z}}^{(i,k)})$ as a function of the sources parameters and the number of side information measurements m_2.

In particular, we obtain sufficient conditions for the MMSE phase transition on the number of projections m_1 and m_2 by considering a (possibly suboptimal) measurement matrix design for $\mathbf{\Phi}_2$, which is inspired on the design which achieves the optimal phase transition for Gaussian sources. In particular, we define $\bar{\mathbf{\Phi}}_2^{(i,k)}$ as in (4.25). Then, we define $\bar{\mathbf{\Phi}}_2 = [(\bar{\mathbf{\Phi}}_2^{(1,1)})^{\mathrm{T}}, \ldots, (\bar{\mathbf{\Phi}}_2^{(K_1,K_2)})^{\mathrm{T}}]^{\mathrm{T}}$ as the matrix obtained by stacking all the matrices $\bar{\mathbf{\Phi}}_2^{(i,k)}$ for all indexes i and k. Finally, the projection kernel $\mathbf{\Phi}_2$ is obtained by m_2 rows from $\bar{\mathbf{\Phi}}_2$. More specifically, $\mathbf{\Phi}_2$ is obtained by picking $m_2^{(i,k)}$ rows from $\bar{\mathbf{\Phi}}_2^{(i,k)}$ and $\sum_{i,k} m_2^{(i,k)} = m_2$.[1]

We will now determine the value of the rank $r_{\mathbf{z}}^{(i,k)}$ when $\mathbf{\Phi}_2$ is obtained by the construction previously described. On using (B.6) and (B.8), we can write

$$r_{\mathbf{z}}^{(i,k)} = r_{\mathbf{x}_1}^{(i,k)} - \Delta^{(i,k)}(\mathbf{\Phi}_2), \tag{B.29}$$

where

$$\begin{aligned}
&\Delta^{(i,k)}(\mathbf{\Phi}_2)\\
&= \mathrm{rank}(\mathbf{\Phi}_2\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)}\mathbf{\Phi}_2^{\mathrm{T}}) - \mathrm{rank}(\mathbf{\Phi}_2(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger}\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})\mathbf{\Phi}_2^{\mathrm{T}}) \qquad \text{(B.30)}\\
&= \dim(\mathrm{Im}(\mathbf{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger}\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}))\\
&\quad - \dim(\mathrm{Im}(\mathbf{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)})). \qquad \text{(B.31)}
\end{aligned}$$

Then, by leveraging the expression of the dimension of the intersection of two linear spaces, we can write

$$\begin{aligned}
&\dim(\mathrm{Im}(\mathbf{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger}\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)}))\\
&= \mathrm{rank}(\mathbf{\Phi}_2) + \mathrm{rank}(\mathbf{N}_{1s}^{(i,k)}) - \mathrm{rank}[\mathbf{\Phi}_2^{\mathrm{T}}\mathbf{N}_{1s}^{(i,k)}] \qquad \text{(B.32)}\\
&= m_2 + n_2 - (r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}) - \mathrm{rank}[\mathbf{\Phi}_2^{\mathrm{T}}\mathbf{N}_{1s}^{(i,k)}], \qquad \text{(B.33)}
\end{aligned}$$

where the columns of $\mathbf{N}_{1s}^{(i,k)} \in \mathbb{R}^{n_2 \times (n_2 - (r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}))}$ form a basis of the linear space $\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger}\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})$. Consider now the last term in (B.33). We can note that

$$\mathrm{rank}[\mathbf{\Phi}_2^{\mathrm{T}}\mathbf{N}_{1s}^{(i,k)}] = \mathrm{rank}[(\hat{\mathbf{\Phi}}_2^{(i,k)})^{\mathrm{T}}\mathbf{N}_{1s}^{(i,k)}], \tag{B.34}$$

where $\hat{\mathbf{\Phi}}_2^{(i,k)}$ is obtained from $\mathbf{\Phi}_2$ by removing the $m_2^{(i,k)}$ rows which are also

[1] In the following, we will assume $m_2 \leq n_2$, since the optimal phase transition for the case $m_2 > n_2$ is trivially obtained by imposing, for example, $\mathbf{\Phi}_2 = [\mathbf{I}_{n_2}\ \mathbf{0}]$.

rows of the matrix $\bar{\mathbf{\Phi}}_2^{(i,k)}$, since all rows in $\bar{\mathbf{\Phi}}_2^{(i,k)}$ are contained in $\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger}\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})$. Moreover, on leveraging the assumption that the linear spaces associated to the images of the covariance matrices of different Gaussian classes (i,k) are independently drawn from a continuous distribution over the Grassmann manifold, we also have

$$\mathrm{rank}[\mathbf{\Phi}_2^{\mathrm{T}}\mathbf{N}_{1s}^{(i,k)}] = \min\{n_2, m_2 - m_2^{(i,k)} + n_2 - (r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)})\}. \tag{B.35}$$

By following similar steps, we can also observe that

$$\dim(\mathrm{Im}(\mathbf{\Phi}_2^{\mathrm{T}}) \cap \mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)})) = m_2 + n_2 - r_{\mathbf{x}_2}^{(i,k)} - \mathrm{rank}[\mathbf{\Phi}_2^{\mathrm{T}}\mathbf{N}_2^{(i,k)}], \tag{B.36}$$

where the column of the matrix $\mathbf{N}_2^{(i,k)} \in \mathbb{R}^{n_2 \times (n_2 - r_{\mathbf{x}_2}^{(i,k)})}$ form a basis of $\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)})$. Then, from the definition of $\bar{\mathbf{\Phi}}_2^{(i,k)}$ in (4.25), we can observe that the rows of $\bar{\mathbf{\Phi}}_2^{(i,k)}$ span a linear space which is contained in $\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)} - \mathbf{\Sigma}_{\mathbf{x}_{21}}^{(i,k)}(\mathbf{\Sigma}_{\mathbf{x}_1}^{(i,k)})^{\dagger}\mathbf{\Sigma}_{\mathbf{x}_{12}}^{(i,k)})$ but which has zero-dimensional intersection with $\mathrm{Null}(\mathbf{\Sigma}_{\mathbf{x}_2}^{(i,k)})$. Moreover, on leveraging the assumption that all the linear spaces associated to the images of the covariance matrices of different Gaussian classes (i,k) are randomly drawn from a continuous distribution over the Grassmann manifold, we also have

$$\mathrm{rank}[\mathbf{\Phi}_2^{\mathrm{T}}\mathbf{N}_2^{(i,k)}] = \min\{n_2, m_2 + n_2 - r_{\mathbf{x}_2}^{(i,k)}\}. \tag{B.37}$$

Finally, on putting together (B.29), (B.31), (B.33), (B.35), (B.36) and (B.37), we can write

$$r_{\mathbf{z}}^{(i,k)} = \min\{r_{\mathbf{x}}^{(i,k)} - m_2, r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)}\} + \max\{m_2 - r_{\mathbf{x}_2}^{(i,k)}, 0\}.$$

Therefore, on considering separately the cases when $m_2 \le r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)}$, $r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \le r_{\mathbf{x}_2}^{(i,k)}$ and $m_2 > r_{\mathbf{x}_2}^{(i,k)}$, we have

$$r_{\mathbf{z}}^{(i,k)} = \begin{cases} r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)} & \text{, if } m_2 \leq r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} \\ \min\{r_{\mathbf{x}}^{(i,k)} - m_2, r_{\mathbf{x}_1}^{(i,k)} - m_2^{(i,k)}\} & \text{, if } r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \leq r_{\mathbf{x}_2}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} & \text{, if } m_2 > r_{\mathbf{x}_2}^{(i,k)} \end{cases} \quad \text{(B.38)}$$

which concludes the proof.

B.4 Proof of Corollary 29

Note that, given any value of m_2, the sufficient conditions for the MMSE to approach zero in the low-noise regime when both $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are random reported in [81, Theorem 4] can be also expressed as

$$m_1 > \begin{cases} r_{\mathbf{x}_1}^{(i,k)} & \text{, if } m_2 \leq r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - m_2 & \text{, if } r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_1}^{(i,k)} < m_2 \leq r_{\mathbf{x}_2}^{(i,k)} \\ r_{\mathbf{x}}^{(i,k)} - r_{\mathbf{x}_2}^{(i,k)} & \text{, if } m_2 > r_{\mathbf{x}_2}^{(i,k)} \end{cases} \quad \text{(B.39)}$$

for all (i, k). Then, the proofs of Corollary 29 is obtained by comparing (B.39) with (4.17), and recalling that $m_2^{(i,k)} \geq 0, \forall (i, k)$.

Chapter 5

A Case Study: Pan-sharpening Imaging Application of RGB with Hyperspectral Images

In this chapter, we present an example of how the linear projection design scheme inspired by the analysis in Chapter 4 can be leveraged in order to improve the performance of real world imaging applications. We first briefly introduce the basic concepts of pan-sharpening and hyperspectral imaging technology and their corresponding applications. We then present our proposed model in Chapter 4 applied to the pan-sharpening application in the presence of compressive hyperspectral data. Finally, our experiment results are also presented to showcase the impact of projection design on applications with real imaging data.

5.1 Introduction

Pan-sharpening, also commonly known as panchromatic sharpening, is an image fusion process that involves merging high-resolution panchromatic and lower resolution multispectral, in order to essentially increase the spatial resolution of a multispectral image [20]. In particular, this technique produces a high-resolution color image from three, four or more low-resolution multispectral satellite bands with one corresponding high-resolution grayscale band. Such band combinations are bundled extracted from a variety of satellite sensors, e.g., QuickBird, IKONOS, OrbView3, Landsat 7. In addition to this, SPOT, GeoEye and DigitalGlobe commercial data packages provide both several colorized low-resolution multispectral

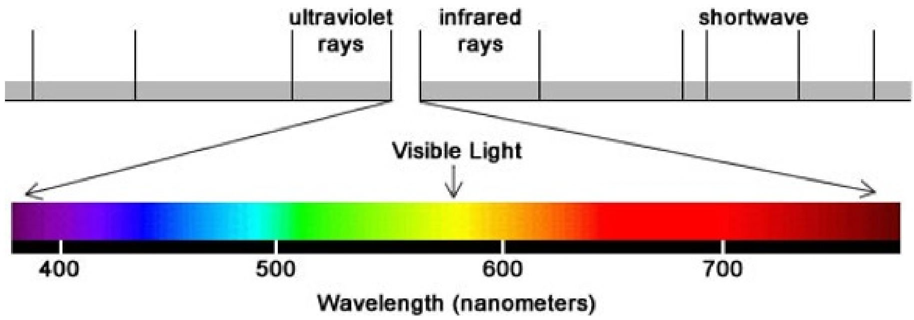


Figure 5.1: Range of wavelengths of visible light [104].

bands and a single high-resolution gray scale band [22].

Multispectral image is a collection of several monochrome images of the same scene, which usually consists of more than one band within specific wavelength ranges across the electromagnetic spectrum [104] as depicted in Fig. 5.1. For example, Landsat-8 produces 11 bands covering from 0.45 to 12.5 μm, providing visible spectrum and as well as the near-infrared band. Each band has a explicit spatial resolution, i.e., Band 10 (Thermal Infrared ($10.60 - 11.19$ μm)) has a spatial resolution of 100 meters. We also note that there is no $0.88-1.36$ μm band due to the atmospheric absorption such that there are no sensors detecting these wavelengths. On the contrary, panchromatic image contains only one wide band of reflectance data, typically spanning a part of the visible part of the spectrum. The range of wavelengths are sampled by imaging detectors, so that panchromatic image is generally collected with higher spatial resolution than a multispectral image because the broad spectral range allows smaller detectors to be used while maintaining a high Signal-to-noise ratio (SNR).

Figs. 5.2(a) and (b) depict the high-resolution panchromatic and low-resolution multispectral images produced by QuickBird satellite [105]. The gray scale image is a scene with 0.6 meter ground resolution, and the low-resolution image represents the same area as shown in the panchromatic image with 2.43 meter ground resolution. Finally, that is the result of fusing a high-resolution panchromatic and a low-resolution multi-spectral image, in which this enhanced high-quality image preserves the original color fidelity and improves better visualization (see Fig. 5.2(c)) [105].

On the other hand, hyperspectral imaging collects and processes informa-

(a) QuickBird panchromatic image, 0.6 meter ground resolution.

(b) QuickBird multispectral image, 2.43 meter ground resolution.

(c) Pan sharpening example. Result has 0.6 meter spatial resolution and 4 multispectral bands.

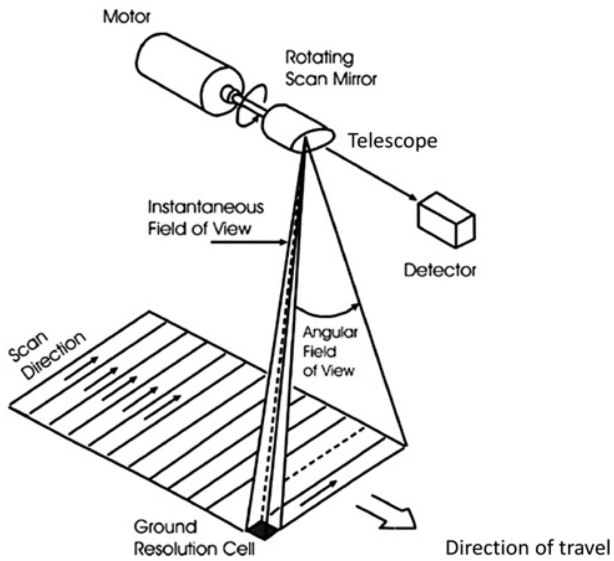
Figure 5.2: QuickBird images and pan-sharpening results [105].

tion from across a wide range of the electromagnetic spectrum, combining the power of digital imaging and spectroscopy [104]. In comparison with multispectral images, hyperspectral images consists of much narrow spectral bands (10-20 nm) over a continuous spectral range, producing the spectrum of all pixels in the scene, where each pixel contains a sampled spectral measurement of reflectance, that can characterize the objects with great precision and detail thus allowing to identify different materials that appear in the scene. In particular, the so called 'images' are combined to form a three-dimensional hyperspectral datacube corresponding to different wavelengths, i.e., a hyperspectral image is a 3D image comprised of a collection of 2D slices, where these 2D slices are captured at different wavelengths. Due to the rich information content in hyperspectral imaging, such imaging technique has been explored for various applications including medical imaging, environmental remote sensing, and astronomy, etc.

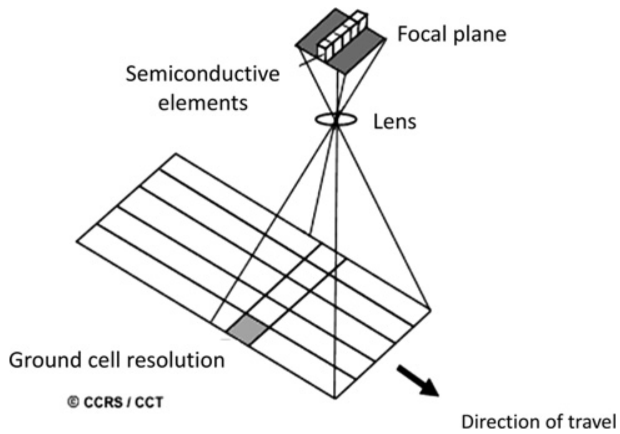
Moreover, there are some basic techniques for collecting the dataset of a hyperspectral data such as whisk broom scanners, push broom scanners [106]. Fig. 5.3 depicts the operating process of these two scanners [107]. A whisk broom scanner is a technology for acquiring satellite images with optical cameras. In whisk broom sensors (commonly known as across track scanner), the device adopts a mirror reflecting light into a single detector in order to collect one pixel of data at a time. Landsat series of satellites use the whisk broom design. Push broom scanner (also known as an along-track scanner) captures images with spectroscopic sensors. Such images are collected with one focal plane array (FPA) measurement per spatial line of the scene (one for each column of pixels). Linear imaging self scanning (LISS) and wide field sensor (WiFS) of Indian remote sensing (IRS) Series and high resolution visible (HRV) of SPOT-1 are the examples of spacecraft cameras using push broom scanners.

However, these acquisition methods cause some disadvantages. For example, data acquisition takes a long time due to the fact that they require scanning a number of zones linearly proportional to a proper spatial and spectral resolution. These conventional spectral imaging techniques can be essentially solved based on compressive sensing paradigm for many spectral imager sampling schemes. The coded aperture snapshot spectral imaging (CASSI) system is commonly used as an example of CS hyperspectral imaging systems, that offers the means to simultaneously sense and capture image data from different wavelengths, in order to reduce the complexity at the expense of capturing such a spatio-spectral data cube. In this sense, CASSI acquires far fewer measurements than those acquired by conventional spectral imagers, thus accelerating the acquisition process.

The CASSI system encodes both 2D spatial and spectral information of ob-



(a) Whisk broom scanner.



(b) Push broom scanner.

Figure 5.3: Whisk broom and push broom scanners [106].

jects using a snapshot from a 3D spatio-spectral information of a scene of interest. More precisely, the CASSI instrument utilizes a coded aperture and one or more dispersive elements to modulate the optical field from a scene. After that, the coded and shifted measurements are detected integrated on a 2D FPA. The prototype of the experimental setup is depicted in Fig. 5.4 [108].

In our case study, we consider a pan-sharpening application, which consists in recovering a high-resolution, color image from a high-resolution panchromatic snapshot and low-resolution hyperspectral images of the same scene. This problem arises often in the context of remote sensing applications [109], where panchromatic sensors are typically used to produce high-resolution images, and hyper-

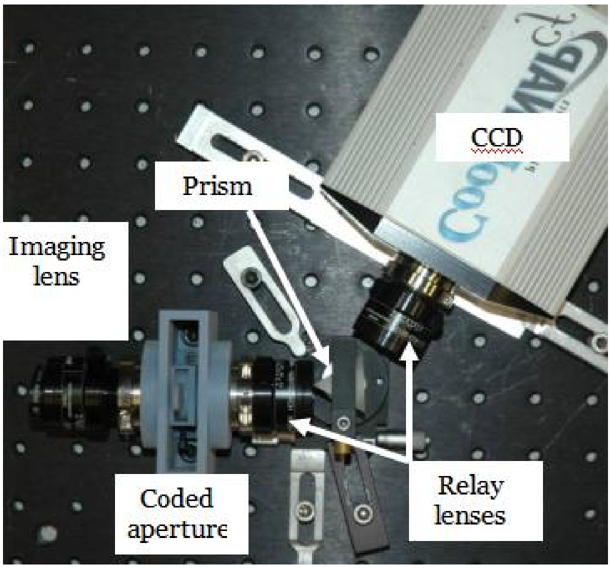


Figure 5.4: The experimental prototype of the system [108].

spectral cameras are used to provide lower resolution images. The use of spatial information from the high-resolution panchromatic image and color information from the low-resolution hyperspectral snapshots can then be combined in order to generate a high-resolution RGB image of the scene. Our case study involves the use of a compressive hyperspectral camera (the CASSI in [110]) – in *lieu* of a standard hyperspectral camera – that provides for a mechanism to design the hyperspectral measurements in order to improve performance further.

5.2 System Model

In our example we propose to replace standard hyperspectral cameras with a compressive CASSI setup [110]. In this scenario, we will show the potential benefits that can be obtained by properly designing the measurements associated to the CASSI architecture in order to optimize panchromatic sharpening, i.e., in order to guarantee reliable recovery of a high-resolution, RGB version of the image of interest from high-resolution panchromatic measurements and low-resolution, compressive hyperspectral measurements. We will see if the our proposed designing principles in previous sections can lead to considerable reconstruction gains.

We consider a high-resolution (512×512 pixels), RGB image of a scene as the signal of interest (Fig. 5.5(a)), and a lower resolution version (256×256 pixels) of a hyperspectral datacube ($N_\lambda = 33$ different channels) of the same scene as the side information signal (Fig. 5.5(b)).

In this case, vectors $\mathbf{x}_1$ represent 8×8 patches extracted from the signal of interest and vectors $\mathbf{x}_2$ represent 4×4 patches corresponding to the same spatial location extracted from the hyperspectral image datacube. In particular, vectors extracted from the signal of interest and the side information are subdivided into overlapping patches with overlap stride equal to 4 and 2 pixels[1], respectively.

The vector $\mathbf{y}_1$ constains high-resolution, panchromatic measurements associated with patches of the image of interest (an example of panchromatic measurements is reported in Fig. 5.6(a)). The linear projection kernel $\mathbf{\Phi}_1$ models the conversion from high-resolution RGB images to a gray scale, panchromatic, high-resolution version of the same image. In particular, on neglecting the effect of gamma compression and expansion[2], and by arranging the RGB channels of the image patches in the vectors $\mathbf{x}_1^R, \mathbf{x}_1^G$, and $\mathbf{x}_1^B$, so that,

$$\mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_1^R \\ \mathbf{x}_1^G \\ \mathbf{x}_1^B \end{bmatrix}, \tag{5.1}$$

we can express $\mathbf{\Phi}_1$ as

$$\mathbf{\Phi}_1 = \begin{bmatrix} c_r\mathbf{I} & c_g\mathbf{I} & c_b\mathbf{I} \end{bmatrix}. \tag{5.2}$$

Here, c_r, c_g, c_b are constant values that are determined by the measured intensity perception of human eyes associated to the three RGB channels. Typical values of such constants are $c_r = 0.2126, c_g = 0.7152, c_b = 0.0722$.

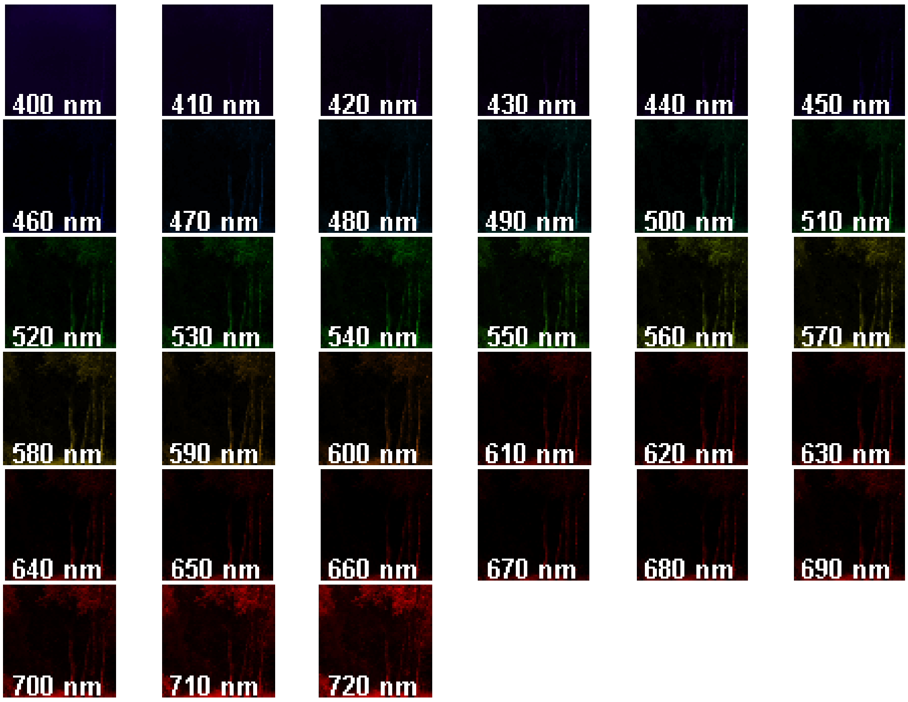
5.2.1 CASSI System

We now consider the vector $\mathbf{y}_2$ contains the side information measurements captured by the compressive hyperspectral imaging device, i.e., the coded aperture snapshot spectral imaging (CASSI) camera (an example of CASSI measurements is reported in Fig. 5.6(b)). Such measurements are based on spectrally shifting the 3D spatio-spectral information (i.e., hyperspectral datacube), through a coded

[1]The overlap stride denotes the distance between corresponding pixel locations in adjacent image patches.

[2]Gamma compression and expansion are one-to-one, nonlinear transformation of the RGB components in each pixel, hence they can be ignored in our measurement model [111].

(a) Original RGB image.



(b) Hyperspectral images.

Figure 5.5: (a) Original RGB image, (b) hyperspectral images corresponding to different wavelengths.

aperture and a dispersive element onto a 2D detector. More precisely, a scene is imaged through imaging optics onto a binary-value coded aperture over all different wavelengths and a dispersive element [112, 113, 114] then projects these multiplex signals of the coded scene with wavelength-dependent shifts (Fig. 5.7 [110]).[3]

[3]For more details on the actual implementation of the CASSI camera, please refer to [110].

(a) Gray scale image.

(b) CASSI measurement.

Figure 5.6: (a) Gray scale image, (b) CASSI measurement.

Then, on writing the vectors $\mathbf{x}_2$, which contain the data in the hyperspectral datacube corresponding to patches, as

$$\mathbf{x}_2 = \begin{bmatrix} \mathbf{x}_2^1 \\ \vdots \\ \mathbf{x}_2^{N_\lambda} \end{bmatrix}, \tag{5.3}$$

where $\mathbf{x}_2^n$ contains the pixels associated to the n-th wavelength channel, then the corresponding CASSI measurement kernel $\mathbf{\Phi}_2$ can be written as

$$\mathbf{\Phi}_2 = \left[\mathbf{M}^1, \ldots, \mathbf{M}^{N_\lambda}\right]. \tag{5.4}$$

Here, the matrices $\mathbf{M}^1, \ldots, \mathbf{M}^{N_\lambda}$ are diagonal with binary diagonal entries that take values 0 and 1 according to the corresponding pixel values in the coding mask.

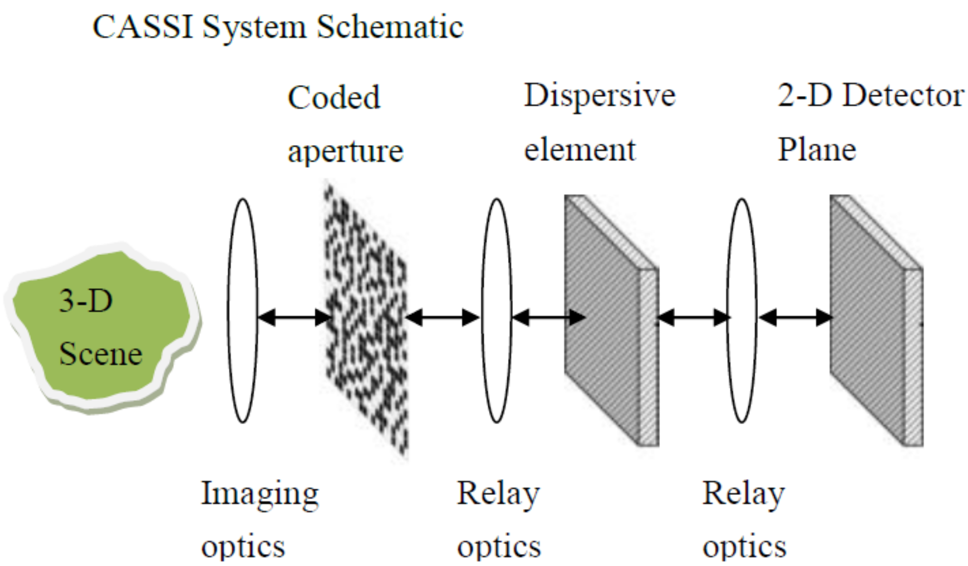


Figure 5.7: Schematic of CASSI.

It is important to note that, since the effects of the coding mask for different wavelengths are obtained by translations of the coded aperture in the device, the elements in the diagonal entries of the different matrices $\mathbf{M}^1, \ldots, \mathbf{M}^{N_\lambda}$ are related one to the other, so it is not possible to choose them independently. In particular, in order to increase the diversity of compressive hyperspectral measurements, and in order to decrease the compression ratio, several snapshots of the same scene can be captured with the CASSI device, by changing the coding mask used for each snapshot. This means that that the design of the measurement matrix $\mathbf{\Phi}_2$ for the CASSI architecture involves the choice of the coded masks pixel values, i.e., the diagonal entries of the matrices $\mathbf{M}^1, \ldots, \mathbf{M}^{N_\lambda}$.

5.3 Experiment Results

We present pan-sharpening reconstruction when the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are modeled by a joint GMM with $K_1 = K_2 = 20$ classes. The parameters of the corresponding distributions, i.e., prior class probabilities, class-conditioned means and class-conditioned covariance matrices, are learnt by using the EM algorithm on the training set described in [110], which contains RGB images and hyperspectral snapshots of 6 different scenes, where 33 hyperspectral images corresponds to 33 different wavelengths ($N_\lambda = 33$) from 400 nm to 720 nm (the training scenes are shown in Fig. 5.8).

We also report reconstruction MMSE results corresponding to two different measurement settings. In the first case, binary coding mask values for CASSI measurements are random, in particular, they are drawn from i.i.d. Bernoulli random variables with probability 0.5. In the second case, the elements of the coding mask are designed according to the construction described in Chapter 4.

However, in view of the fact that the CASSI architecture only permits one to constitute a certain class of coded apertures (see (5.4)) – our design is obtained as follows:

$$\mathbf{\Phi}_2^{\mathrm{CASSI}} = \underset{\mathbf{\Phi}_2 \in \mathcal{F}_{\mathrm{CASSI}}}{\operatorname{argmin}} \; ||\mathbf{\Phi}_2 - \mathbf{\Phi}_2^{\mathrm{GSVD}}||_F \tag{5.5}$$

where $\mathbf{\Phi}_2^{\mathrm{CASSI}}$ embodies the CASSI coded aperture design, $\mathbf{\Phi}_2^{\mathrm{GSVD}}$ embodies the design in (4.25), and $\mathcal{F}_{\mathrm{CASSI}}$ denotes the set of coded aperture designs consistent with the CASSI architecture. That is, the proposed design $\mathbf{\Phi}_2^{\mathrm{CASSI}}$ is the closest in Frobenius norm to the GSVD design $\mathbf{\Phi}_2^{\mathrm{GSVD}}$ in (4.25), that is consistent with the CASSI constraints.

The diagonal elements of the matrices $\mathbf{M}^1, \ldots, \mathbf{M}^{N_\lambda}$ in $\mathbf{\Phi}_2$ are generated by shifting the binary pattern in the coding mask. In particular, on considering single pixel shifts in a given direction, patches of size 4×4 and $N_\lambda = 33$ channels, the total number of pixels used to code a given patch is $4 \times (4 + 33 - 1) = 144$. We note that a brute-force approach in solving the optimization problem (5.5) would require the computation of 2^{144} possible matrices in the considered scenario. In fact, the optimization problem (5.5) can be solved with computational complexity growing linearly with the number of binary pixel values corresponding to the coding mask. In this way, we design possible 0s and 1s for the coding mask by minimizing the distance between the corresponding non-zero entries in the $\mathbf{\Phi}_2$ that implementing CASSI measurements, and the corresponding elements in the projection kernel $\mathbf{\Phi}_2^{\mathrm{GSVD}}$. Moreover, we use the same mask for all patches contained in the image of interest.

Fig. 5.9 reports reconstruction examples of the RGB image in Fig. 5.5(a), and the corresponding PSNR values are reported in Table 5.1. Figs. 5.9(a) and (b) show reconstructions obtained with random and designed CASSI measurements, respectively, obtained with a single snapshot[4]. We can observe that designed CASSI measurements can guarantee significant improvement with respect to random measurements. In particular, designed measurements are particularly effective in reducing the red color artifacts which are easily observable in Fig. 5.9(a). A similar effect is also observed Figs. 5.9(d) and (e) which contain reconstructions examples for random and designed measurements corresponding to two CASSI snapshots. In fact, as reported in Table 5.1, for both cases of one and two CASSI snapshots, the proposed measurement design scheme leads to a PSNR gain of approximately 2 dB. In this experiment, we also provide the reconstruction results by using designed linear measurements to capture the hyperspectral images, i.e., the projection kernel $\mathbf{\Phi}_2^{\mathrm{GSVD}}$ embodies the design in (4.25). Note that the designed linear measurements do not follow the CASSI architecture. As reported in Table 5.1 and Figs. 5.9(c) and (f), we observe that the case of designed linear measurements can lead to reconstruction PSNR gains of about 3 dB in relation to the case of designed CASSI measurements.

Overall, these gains are particularly significant by taking into account that the optimization of the measurement procedure has been effected under the strong constraints imposed by the CASSI architecture. Therefore, this also leads us to conclude that (4.25) or (5.5) for GMM sources can be used to guide how to capture side information in compressive sensing systems.

We now consider the scenario when all parameters of the joint GMM are with

[4]In this case, since $n_2 = 4 \times 4 \times 33 = 528$, matrices $\mathbf{\Phi}_2$ corresponding to a single CASSI snapshot have dimensions 16×528, i.e., $m_2 = 16$, whereas matrices corresponding to two snapshots have dimensions 32×528, i.e., $m_2 = 32$.

Table 5.1: PSNR values of the reconstruction examples in Fig. 5.9 with $K_1 = K_2 = 20$.

	Random CASSI	Designed CASSI	Designed GSVD
1 snapshot	25.4633 dB	28.2043 dB	31.7402 dB
2 snapshots	28.5474 dB	30.6788 dB	32.8802 dB
4 snapshots	28.7205 dB	30.8428 dB	33.4992 dB
6 snapshots	28.8172 dB	31.1200 dB	33.8355 dB
8 snapshots	29.1690 dB	31.8641 dB	34.0250 dB
10 snapshots	29.3645 dB	31.8980 dB	34.1162 dB

Table 5.2: PSNR values of the reconstruction examples in Fig. 5.10 with $K_1 = K_2 = 10$.

	Random CASSI	Designed CASSI	GSVD design
1 snapshot	24.6691 dB	26.9297 dB	30.4790 dB
2 snapshots	26.7197 dB	30.6380 dB	31.9180 dB
4 snapshots	27.7300 dB	30.6643 dB	32.4233 dB
6 snapshots	28.2733 dB	30.9505 dB	32.9705 dB
8 snapshots	28.7604 dB	31.6680 dB	33.5762 dB
10 snapshots	29.1008 dB	31.8092 dB	33.9034 dB

$K_1 = K_2 = 10$. In this case, we also consider vectors $\mathbf{x}_1$ represent 8×8 patches extracted from the signal of interest and vectors $\mathbf{x}_2$ represent 4×4 patches corresponding to the same spatial location extracted from the hyperspectral image datacube. All parameters of the corresponding distributions, i.e., prior class probabilities, class-conditioned means and class-conditioned covariance matrices, are trained by using the EM algorithm on the training set described in [110], which contains RGB images and hyperspectral snapshots of 6 different scenes, where 33 hyperspectral images corresponds to 33 different wavelengths ($N_\lambda = 33$) from 400 nm to 720 nm (the training scenes are shown in Fig. 5.8). In this experiment, we consider the projection matrix $\mathbf{\Phi}_2$ is designed in (5.5), and we also consider CASSI measurements are randomly drawn from i.i.d. Bernoulli variables with probability 0.5 as comparison purpose.

Fig. 5.10 reports reconstruction examples of the RGB image in Fig. 5.5(a). The results in Figs. 5.10(a) and (b) show reconstructions obtained with random and designed CASSI measurements, respectively, by considering a single CASSI snapshot. The results for optimal CASSI measurements exhibit PSNR gain of approximately 2 dB with respect to the case of random projection kernels. As in previous experiment, designed measurements are particularly effective in reducing the red color artifacts. We also consider two CASSI snapshots as the side information in order to lead to better reconstruction performance as shown in Figs. 5.10(d) and (e). In fact, as reported in Table 5.2, for both cases of one and two CASSI snapshots, the proposed measurement design scheme leads to a PSNR gain of approximately 2 dB. Furthermore, we also consider the scenario when the projection kernel associated with the hyperspectral images is linear, i.e. the projection kernel $\mathbf{\Phi}_2^{\mathrm{GSVD}}$ is generated in (4.25) which does not follow the CASSI architecture. As reported in Table 5.2 and and Figs. 5.10(c) and (f), the case of designed linear measurements can lead to reconstruction PSNR gains in excess of 3 dB over the case of designed CASSI measurements. In this experiment, we can also observe that patches described by a 20-classes joint GMM distribution provides better reconstruction performance than the case where patches described by a 10-classes joint GMM distribution.

Finally, we consider the scenario when all parameters of the joint GMM are with $K_1 = K_2 = 30$. We also use a high-resolution RGB image (512×512 pixels) as the signal of interest, and a lower resolution version (256×256 pixels) of a hyperspectral datacube ($N_\lambda = 33$ different channels) as the side information. In this case, we consider both signals are partitioned into patches, so that vectors $\mathbf{x}_1$ represent 8×8 patches extracted from the signal of interest and vectors $\mathbf{x}_2$ represent 4×4 patches corresponding to the same spatial location extracted from the hyperspectral image datacube. As in previous experiment, all parameters of the corresponding distributions, i.e., prior class probabilities, class-conditioned means and class-conditioned covariance matrices, are trained by using the EM algorithm on the training set described in [110], which contains original RGB and hyperspectral images of 6 different scenes, where 33 hyperspectral images corresponds to 33 different wavelengths ($N_\lambda = 33$) from 400 nm to 720 nm. In this experiment, we consider the projection matrix $\mathbf{\Phi}_2$ is designed in (5.5), and we also consider CASSI measurements are randomly drawn from i.i.d. Bernoulli variables with probability 0.5 as comparison purpose.

The results in Fig. 5.11 report reconstruction examples of the RGB image in Fig. 5.5(a). In this experiment, Figs. 5.11(a) and (b) show reconstructions

Table 5.3: PSNR values of the reconstruction examples in Fig. 5.10 with $K_1 = K_2 = 30$.

	Random CASSI	Designed CASSI	GSVD design
1 snapshot	25.9441 dB	28.7297 dB	31.8320 dB
2 snapshots	28.9102 dB	31.1056 dB	32.6602 dB
4 snapshots	29.1105 dB	31.4022 dB	33.6415 dB
6 snapshots	29.5082 dB	31.5980 dB	34.1850 dB
8 snapshots	29.8874 dB	31.7464 dB	34.2933 dB
10 snapshots	30.1010 dB	31.9541 dB	34.3010 dB

obtained with random and designed CASSI measurements, respectively, by considering a single CASSI snapshot. In particular, designed measurements are particularly effective in reducing the red color artifacts as in previous experiments. In turn, Figs. 5.11(d) and (e) contain reconstructions examples for random and designed measurements corresponding to two CASSI snapshots depict similar effects. In fact, as reported in Table 5.3, for both cases of one and two CASSI snapshots, the proposed measurement design scheme improves reconstruction performance in terms of PSNR gain of approximately 2 dB. In this experiment, we also consider the scenario when the projection kernel associated with the hyperspectral images is linearly designed, i.e. the projection kernel $\mathbf{\Phi}_2^{\text{GSVD}}$ embodies the design in (4.25). We can observe that the design in (4.25) can lead to PSNR gains of approximately 3 dB in relation to the design in (5.5), as reported in Table 5.3 and Figs. 5.11(c) and (f). Finally, we can also observe that patches described by a higher joint GMM distribution provides better reconstruction performance.

We also compare our GMM approach to linking the two different data modalities, i.e. the signal of interest and the side information signal, to another approach based on sparse representations induced by coupled dictionary learning (CDL) [74, 115] (note that this approach to linking different data modalities generalizes the JSM-1 and JSM-3 models proposed in [83] and [84]). In particular, we consider dictionaries with 512 atoms and common and innovation components with sparsities equal to 4 and 2, respectively. We use the training data shown in Fig. 5.8 to learn dictionaries that link the hyperspectral patches and the RGB patches using the method described in [74, 115]. The dimensions and strides of the patches used for dictionary learning are the same as those used in the previous experiment.

Table 5.4: PSNR values of the reconstruction examples by using the CDL approach.

	CDL approach
1 snapshot	26.1185 dB
2 snapshots	28.0251 dB
4 snapshots	28.3192 dB
6 snapshots	28.4308 dB
8 snapshots	29.2106 dB
10 snapshots	29.5611 dB

We then use the testing data (as shown in Fig. 5.5(a)) to determine the reconstruction performance. In particular, we use the orthogonal matching pursuit (OMP) algorithm to reconstruct the signal of interest from noisy random linear projections of the signal of interest as well as noisy random linear projections from the side information. Here, note that we consider random measurements (as described earlier) rather than designed ones because it is not immediate how to conceive optimized linear measurement designs to capture the side information for this approach to linking the different data modalities. We note again that our proposed design leads to PSNR gains of about 2 dB in relation to the CDL approach (see Tables 5.1 and 5.4).

Finally, we present PSNR values with different number of snapshots by using different sensing approaches, as depicted in Fig. 5.12. In particular, we consider a joint GMM with $K_1 = K_2 = 20$ for the cases of random CASSI, designed CASSI and designed GSVD measurements. We have shown that our linear measurement designs deliver better reconstruction results in relation to random ones, both when one couples the different data modalities using a GMM model or else a CDL model. Of particular relevance, one of the advantages of using GMM models in relation to the CDL ones is that one can come up with concrete strategies to define optimal measurement designs.

(a) Training scene 1.

(b) Training scene 2.

(c) Training scene 3.

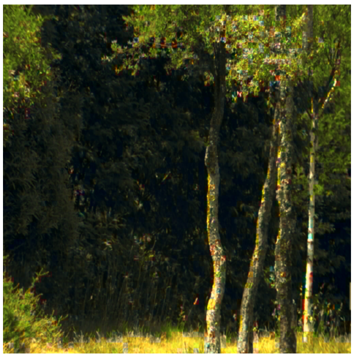
(d) Training scene 4.

(e) Training scene 5.

(f) Training scene 6.

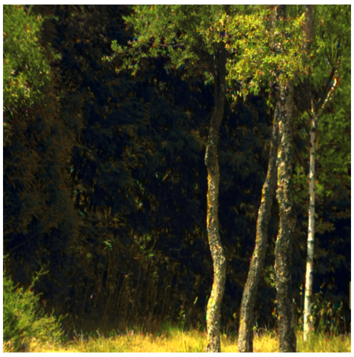
Figure 5.8: Training dataset.

(a) Random CASSI measurements (1 snapshot).



(b) Designed CASSI measurements (1 snapshot).

(c) Designed GSVD measurements (1 snapshot).

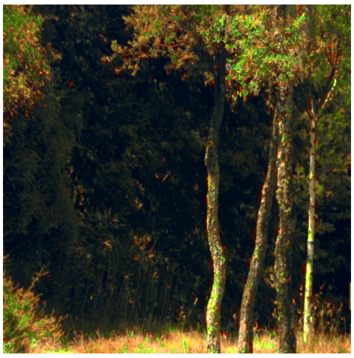


(d) Random CASSI measurements (2 snapshots).

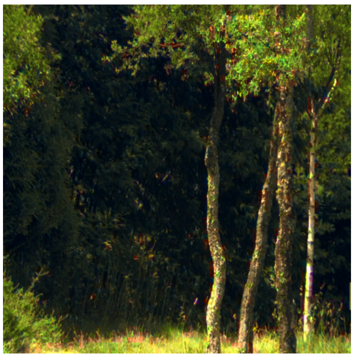
(e) Designed CASSI measurements (2 snapshots).

(f) Designed GSVD measurements (2 snapshot).

Figure 5.9: Reconstruction examples for random and designed CASSI measurements, 1 snapshot ($m_2 = 16$) and 2 snapshots ($m_2 = 32$) with $K_1 = K_2 = 20$.



(a) Random CASSI measurements (1 snapshot).



(b) Designed CASSI measurements (1 snapshot).

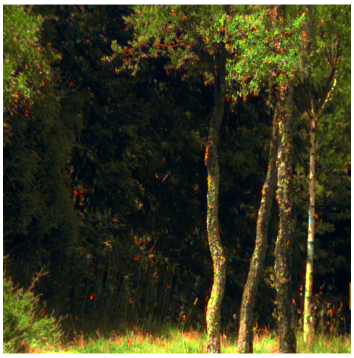
(c) Designed GSVD measurements (1 snapshot).

(d) Random CASSI measurements (2 snapshots).

(e) Designed CASSI measurements (2 snapshots).

(f) Designed GSVD measurements (2 snapshot).

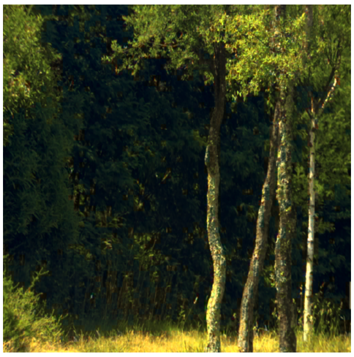
Figure 5.10: Reconstruction examples for random and designed CASSI measurements, 1 snapshot ($m_2 = 16$) and 2 snapshots ($m_2 = 32$) with $K_1 = K_2 = 10$.



(a) Random CASSI measurements (1 snapshot).

(b) Designed CASSI measurements (1 snapshot).

(c) Designed GSVD measurements (1 snapshot).



(d) Random CASSI measurements (2 snapshots).

(e) Designed CASSI measurements (2 snapshots).

(f) Designed GSVD measurements (2 snapshot).

Figure 5.11: Reconstruction examples for random and designed CASSI measurements, 1 snapshot ($m_2 = 16$) and 2 snapshots ($m_2 = 32$) with $K_1 = K_2 = 30$.

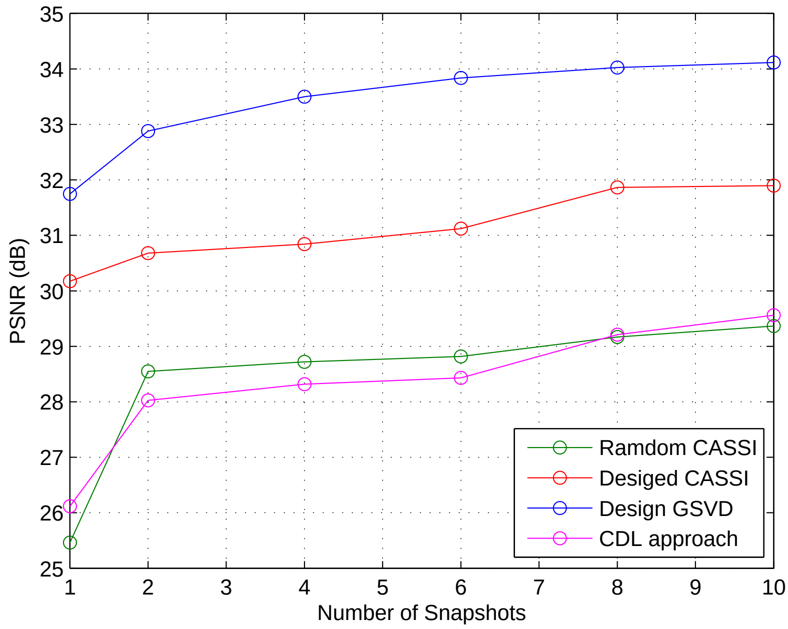


Figure 5.12: PSNR values with different number of snapshots by using four approaches.

(a) Coupled dictionary learning (1 snapshot).

(b) Coupled dictionary learning (2 snapshots).

Figure 5.13: Reconstruction examples for coupled dictionary learning (CDL) based approach, 1 snapshot ($m_2 = 16$) and 2 snapshots ($m_2 = 32$).

5.4 Summary

The case study has showcased that our proposed measurement designs in Chapter 4 can potentially apply to real-world imaging with side information in CS. In particular, we have considered pan-sharpening applications, by reconstructing a high-resolution, color image from a high-resolution panchromatic snapshot and low-resolution hyperspectral images of the same scene. Such measurements extracted from the hyperspectral images are performed with the use of a compressive hyperspectral camera (CASSI).

In this work, the projection kernel associated to the side information implicate strong constraints in designing measurements with the CASSI architecture, by assuming the projection matrix are diagonal with binary diagonal entries that take values 0 and 1. Perhaps surprisingly, under the challenging scenario, we have obtained 2 dB PSNR gains of reconstruction by using closest projection kernel in Frobenius norm to the designed kernel in (4.25), involving the set of coded aperture designs consistent with the CASSI architecture.

Chapter 6

Conclusions

6.1 Summary of the Thesis

This thesis studies how to design linear projection kernels in order to improve reconstruction performance of high-dimensional signals from noisy, linear low-dimensional signal features in the presence of side information. In particular, the focus has been on the design of kernels to optimally capture the signal of interest and the design of kernels to optimally capture the side information signal. The focus has also been on scenarios where both the target signal and the side information signal are drawn from a joint GMM distribution, with possibly low-rank, class-conditioned input covariance matrices.

The main contributions of the work have been:

1. Linear Projection Designs to Capture the Signal of Interest

In Chapter 3, we have developed a framework to study the design of linear projection kernels to capture a target signal in the presence of a side information signal. In particular, this framework encompasses both the scenario where the side information is available to the decoder only and the scenario where the side information is available both to the decoder and the encoder.

Our main contribution is a theoretical characterization of the minimum number of measurements needed to guarantee the reconstruction MMSE to approach zero in both scenarios. Of particular relevance, our theoretical characterizations show that such minimum number of measurements is the same both in the scenario where the side information is available only to decoder and scenario where the side information is available both to decoder and encoder.

We have also provided numerical results for synthetic data drawn from both Gaussian and GMM distributions in order to confirm our proposed theorems are well aligned with practice. In particular, the MMSE results for both design cases have illustrated the phase transition with designed kernel is the same obtained for random projection kernels. Nevertheless, we have observed significant gains of the reconstruction performance for finite noise levels for both designs. Our design approach has also applied to the real data, where it was shown it can lead to considerable PSNR gains of images reconstruction with respect to the case of random.

2. Linear Projection Designs to Capture the Side Information

In Chapter 4, we have also developed a framework to study the design of linear projection kernels to capture the side information signal.

Our main contributions are: (1) a theoretical characterization of the minimum number of measurements needed to guarantee the target signal reconstruction MMSE to approach zero in the low-noise regime; and (2) a concrete linear projection designs that guarantee the target signal reconstruction MMSE to approach zero in the low-noise regime. In particular, it was shown that such linear projection designs aims to capture the most of signal components from side information that are maximally correlated with the signal of interest. It was also shown that measurements design for side information can lead to substantial gains over random linear designs in terms of the number of measurements needed for the reconstruction MMSE to approach zero in the low-noise regime.

A number of experiments with synthetic data both for Gaussian and GMM sources have confirmed that our theoretical results are very well aligned with numerical ones. We have observed that careful design of the projection kernel for the side information leads to reliable reconstruction with a lower number of measurements with respect to the random one. Numerical results were also presented to showcase the impact of projection design on applications with real imaging data. We have observed that our design approach obtains significant gains of reconstruction performance with respect to the case of random projection kernels.

3. Case Study: Reconstruction of RGB Images from Gray-Scale ones in the Pres-

ence of Compressive Hyperspectral Measurements

Finally, in Chapter 5, this thesis has considered a particular real world imaging application – namely, a pan-sharpening application associated with the reconstruction of a high-resolution, RGB image from a high-resolution panchromatic snapshot and low-resolution hyperspectral images of the same scene – in order to showcase the merit of our proposed linear projection designs. In particular, by considering a scenario where the hyperspectral data was provided by the CASSI architecture, we have shown how to use our proposed linear project designs from Chapter 4 to guide the design of the linear projections for this application. We have also illustrated that our proposed linear projection designs can lead to substantial performance improvements in relation to random ones.

In summary, this thesis concentrates on compressive sensing systems that entail the acquisition and reconstruction of some target signal in the presence of a side information signal. In this context, this thesis has investigated how to capture the target signal as well as the side information signal by providing bounds on the number of measurements required for reliable reconstruction, and by providing concrete linear projection designs. The merit of our approach has been validated via extensive numerical results both with real and synthetic data and via a case study that arises in remote sensing applications.

6.2 Future Work

This thesis opens up various future research directions. These include:

1. Optimal (simultaneous) acquisition of target signal and side information signal

This thesis has concentrated on the linear projection designs to capture either the target signal or else the side information signal (where the target signal is randomly measured). It would also be of interest to study how to optimize simultaneously the linear projection kernel associated with the target signal and the linear projection kernel associated with the side information signal in order to understand whether additional gains are possible.

Moreover, it would also be of interest to generalize the theoretical results from the scenario where one wishes to optimize for the MMSE phase transition to the scenario where one also wishes to optimize additionally the MMSE dimension. In particular, reference [93] has established that the MMSE can obey an expansion akin to $\mathrm{MMSE}(\sigma^2) = \mathcal{M}_\infty + \mathcal{D} \cdot \sigma^2 + o(\sigma^2)$, where $\mathcal{M}_\infty$ is the zero-order term in the expansion, which relates to the MMSE floor, the term $\mathcal{D}$ is the MMSE dimension, and σ^2 is the noise variance. Therefore, a more rigorous projection design approach ought to minimize $\mathcal{M}_\infty$ but also $\mathcal{D}$. Such an approach could reveal additional insights about how to optimize projects at finite noise levels.

2. Matrix recovery, matrix completion, tensor recovery and tensor completion

This thesis has also concentrated on scenarios where one wishes to reconstruct a signal from noisy compressive measurements but there are various other scenarios where one intends instead to reconstruct a matrix or a tensor from compressive measurements [116]. It would also be of interest to study how to generalize the current framework from the scenario where one wishes to reconstruct a signal from its noisy compressive measurements in the presence of side information to the case where one wishes to reconstruct a matrix or a tensor from noisy measurements in the presence of side information. The impact of side information in these class of problems is not yet very well studied.

3. Applications

The focus of the thesis has been on the application of the proposed linear projection designs in a pan-sharpening problem arising in remote sensing applications. However, it would also be interesting to study how to apply our linear projection designs in other application domains, such as in medical imaging.

In particular, the feasibility of using compressive sensing approaches in MRI has been demonstrated in various works due to the fact that MRI are naturally sparse in Fourier (or some frequency) domain [7, 10].

Furthermore, PET/MRI scanners have been available for a few years and have also been discussed in conjunction with CS [70], which would be of interest to study how to optimally capture PET and MR images, in order to obtain reliable quality of image reconstruction with less required measurements thus implying to accelerate the scanning time of imaging process.

Moreover, the common MRI sequences – T1 and T2 weighted scans where these images are produced corresponding to short *time to echo* (TE) and *repeti-*

tion time (TR), and long TE and TR, respectively. It would also be of interest to discuss how to simultaneously integrate MRI with PET by increasing or decreasing parameters of T1 and T2 for measuring different tissue contrast in order to efficiently obtain reliable reconstruction.

Bibliography

[1] B. P. Lathi, *Modern digital and analog communication systems.* Oxford University Press, Inc., 1990.

[2] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[3] R. G. Vaughan, N. L. Scott, and D. R. White, "The theory of bandpass sampling," *IEEE Transactions on signal processing*, vol. 39, no. 9, pp. 1973–1984, 1991.

[4] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[7] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[8] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Radar Conference, 2007 IEEE*, pp. 128–133, IEEE, 2007.

[9] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications.* Cambridge University Press, 2012.

[10] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing mri,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[11] L. Weizman, Y. C. Eldar, and D. Ben Bashat, “Compressed sensing for longitudinal mri: An adaptive-weighted approach,” *Medical physics*, vol. 42, no. 9, pp. 5195–5208, 2015.

[12] X. X. Zhu and R. Bamler, “Tomographic sar inversion by $l\_\{1\}$-norm regularizationthe compressive sensing approach,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, 2010.

[13] W. Wang, B. Zhang, J. Mu, and X. Wu, “Sparsity and compressive sensing for sar signal,” in *Globecom Workshops (GC Wkshps), 2012 IEEE*, pp. 1416–1419, IEEE, 2012.

[14] S. R. Cherry, “Multimodality in vivo imaging systems: twice the power or double the trouble?,” *Annu. Rev. Biomed. Eng.*, vol. 8, pp. 35–62, 2006.

[15] Geosage, “Image fusion and spectral transformation.” `http://www.geosage.com/highview/imagefusion.html`. [Online; accessed 30.06.2017].

[16] P. Song, J. F. Mota, N. Deligiannis, and M. R. Rodrigues, “Coupled dictionary learning for multimodal image super-resolution,” in *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*, pp. 162–166, IEEE, 2016.

[17] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Deep joint image filtering,” in *European Conference on Computer Vision*, pp. 154–169, Springer, 2016.

[18] D. Townsend, “Multimodality imaging of structure and function,” *Physics in medicine and biology*, vol. 53, no. 4, p. R1, 2008.

[19] C. Catana, A. R. Guimaraes, and B. R. Rosen, “Pet and mr imaging: the odd couple or a match made in heaven?,” *Journal of Nuclear Medicine*, vol. 54, no. 5, pp. 815–824, 2013.

[20] Z. Chen, H. Pu, B. Wang, and G.-M. Jiang, "Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1418–1422, 2014.

[21] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7738–7749, 2014.

[22] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.

[23] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[24] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[25] V. A. KOTEL'NIKOV, "On the carrying capacity of the" either" and wire in telecommunications," in *Material for the First All-Union Conference on Questions of Communication (Russian), Izd. Red. Upr. Svyzai RKKA, Moscow, 1933*, 1933.

[26] E. Whitaker, "On the functions which are represented by the expansion of interpolating theory," in *Proc. Roy. Soc. Edinburgh*, vol. 35, pp. 181–194, 1915.

[27] J. M. Whittaker, *Interpolatory function theory*, vol. 33. The University Press, 1935.

[28] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on selected areas in communications*, vol. 17, no. 4, pp. 539–550, 1999.

[29] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[30] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[31] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[32] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[33] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive sensing," *SIAM journal on scientific computing*, vol. 33, no. 1, pp. 250–278, 2011.

[34] M. Schmidt, G. Fung, and R. Rosales, "Fast optimization methods for l1 regularization: A comparative study and two new approaches," in *European Conference on Machine Learning*, pp. 286–297, Springer, 2007.

[35] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[36] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[37] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[38] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[39] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.

[40] M. Chen, J. Silva, J. Paisley, C. Wang, D. Dunson, and L. Carin, "Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6140–6155, 2010.

[41] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of computational mathematics*, vol. 9, no. 1, pp. 51–77, 2009.

[42] C. Chen and J. Huang, "Compressive sensing mri with wavelet tree sparsity," in *Advances in neural information processing systems*, pp. 1115–1123, 2012.

[43] V. M. Patel, G. R. Easley, D. M. Healy, and R. Chellappa, "Compressed sensing for synthetic aperture radar imaging," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 2141–2144, IEEE, 2009.

[44] V. M. Patel, G. R. Easley, D. M. Healy Jr, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE Journal of selected topics in signal processing*, vol. 4, no. 2, pp. 244–254, 2010.

[45] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[46] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[47] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, vol. 1. Birkhäuser Basel, 2013.

[48] M. Fornasier and H. Rauhut, "Compressive sensing," in *Handbook of mathematical methods in imaging*, pp. 187–228, Springer, 2011.

[49] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1248–1259, 2014.

[50] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[51] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[52] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[53] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[54] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, 2011.

[55] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.

[56] R. Giryes and M. Elad, "Rip-based near-oracle performance guarantees for sp, cosamp, and iht," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1465–1468, 2012.

[57] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal match-

ing pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[58] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[59] C. La and M. N. Do, "Tree-based orthogonal matching pursuit algorithm for signal reconstruction," in *Image Processing, 2006 IEEE International Conference on*, pp. 1277–1280, IEEE, 2006.

[60] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on information theory*, vol. 52, no. 1, pp. 6–18, 2006.

[61] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.

[62] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6215–6221, 2011.

[63] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, "One sketch for all: fast algorithms for compressed sensing," in *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pp. 237–246, ACM, 2007.

[64] A. C. Gilbert, S. Muthukrishnan, and M. Strauss, "Improved time bounds for near-optimal sparse fourier representations," in *Proceedings of SPIE*, vol. 5914, p. 59141A, 2005.

[65] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, "Algorithmic linear dimension reduction in the l_1 norm for sparse vectors," *arXiv preprint cs/0608079*, 2006.

[66] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[67] M. Fornasier and H. Rauhut, "Iterative thresholding algorithms," *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, pp. 187–208, 2008.

[68] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.

[69] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 269–280, 2010.

[70] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge, "Joint reconstruction of pet-mri by exploiting structural similarity," *Inverse Problems*, vol. 31, no. 1, p. 015001, 2014.

[71] L. Weizman, Y. C. Eldar, and D. B. Bashat, "The application of compressed sensing for longitudinal mri," tech. rep., 2014.

[72] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets," *Medical physics*, vol. 35, no. 2, pp. 660–663, 2008.

[73] X. Yuan, T.-H. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 964–976, 2015.

[74] N. Deligiannis, J. F. Mota, B. Cornelis, M. R. Rodrigues, and I. Daubechies, "Multi-modal dictionary learning for image separation with application in art investigation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 751–764, 2017.

[75] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.

[76] X. Wang and J. Liang, "Side information-aided compressed sensing reconstruction via approximate message passing," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 3330–3334, IEEE, 2014.

[77] J. F. Mota, N. Deligiannis, and M. R. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," *arXiv preprint arXiv:1408.5250*, 2014.

[78] J. F. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. R. Rodrigues, "Adaptive-rate sparse signal reconstruction with application in compressive background subtraction," *arXiv preprint arXiv:1503.03231*, 2015.

[79] J. F. Mota, L. Weizman, N. Deligiannis, Y. C. Eldar, and M. R. Rodrigues, "Reference-based compressed sensing: A sample complexity approach," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pp. 4687–4691, IEEE, 2016.

[80] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. R. Rodrigues, "Classification and reconstruction of compressed gmm signals with side information," in *Information Theory (ISIT), 2015 IEEE International Symposium on*, pp. 994–998, IEEE, 2015.

[81] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, R. Calderbank, L. Carin, and M. R. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional features in the presence of side information," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6459–6492, 2016.

[82] P. Song, J. F. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Measurement matrix design for compressive sensing with side information at the encoder,"

in *Statistical Signal Processing Workshop (SSP), 2016 IEEE*, pp. 1–5, IEEE, 2016.

[83] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," *arXiv preprint arXiv:0901.3403*, 2009.

[84] M. Duarte, M. Wakin, D. Baron, S. Sarvotham, and R. Baraniuk, "Measurement bounds for sparse signal ensembles via graphical models," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4280–4289, 2013.

[85] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[86] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[87] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Transactions on signal processing*, vol. 46, no. 4, pp. 886–902, 1998.

[88] G. Yu and G. Sapiro, "Statistical compressed sensing of gaussian mixture models," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5842–5858, 2011.

[89] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Task-driven adaptive statistical compressive sensing of gaussian mixture models," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 585–600, 2013.

[90] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, 2012.

[91] M. Chen, W. Carson, M. Rodrigues, R. Calderbank, and L. Carin, "Communications inspired linear discriminant analysis," *arXiv preprint arXiv:1206.6397*, 2012.

[92] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using gaussian mixture models," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4863–4878, 2014.

[93] F. Renna, R. Calderbank, L. Carin, and M. R. Rodrigues, "Reconstruction of signals drawn from a gaussian mixture via noisy compressive measurements," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2265–2277, 2014.

[94] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint tx-rx beamforming design for multicarrier mimo channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, 2003.

[95] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

[96] C. M. Bishop, *Pattern recognition and machine learning.* springer, 2006.

[97] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 594–611, 2006.

[98] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge university press, 2012.

[99] B.-Y. Wang, X. Zhang, and F. Zhang, "Some inequalities on generalized schur complements," *Linear Algebra and its Applications*, vol. 302, pp. 163–172, 1999.

[100] C. C. Paige and M. A. Saunders, "Towards a generalized singular value decomposition," *SIAM Journal on Numerical Analysis*, vol. 18, no. 3, pp. 398–405, 1981.

[101] Y. Tian and G. P. Styan, "Rank equalities for idempotent matrices with applications," *Journal of Computational and Applied Mathematics*, vol. 191, no. 1, pp. 77–97, 2006.

[102] D. V. Ouellette, "Schur complements and statistics," *Linear Algebra and its Applications*, vol. 36, pp. 187–295, 1981.

[103] J. Gallier, "The Schur complement and symmetric positive semidefinite (and definite) matrices," *December*, vol. 44, pp. 1–12, 2010.

[104] Q. Li, X. He, Y. Wang, H. Liu, D. Xu, and F. Guo, "Review of spectral imaging technology in biomedical engineering: achievements and challenges," *Journal of biomedical optics*, vol. 18, no. 10, pp. 100901–100901, 2013.

[105] K. University of Rhode Island, "Create a pansharpened image using arcgis." `https://www.edc.uri.edu/blog/part-3-create-pansharpened-image-using-arcgis`. [Online; accessed 20.06.2017].

[106] P. Shippert, "Push broom and whisk broom sensors," 2013.

[107] D. J. Keith, "Coastal and estuarine waters: Optical sensors and remote sensing for management," 2014.

[108] D. University, "The duke imaging and spectroscopy program." `https://http://www.disp.duke.edu/projects/CASSI/index.ptml`. [Online; accessed 30.06.2017].

[109] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4779–4789, 2013.

[110] A. Rajwade, D. Kittle, T.-H. Tsai, D. Brady, and L. Carin, "Coded hyperspectral imaging and blind compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 782–812, 2013.

[111] C. A. Poynton, "Rehabilitation of gamma," in *Human Vision and Electronic Imaging*, vol. 111, pp. 232–249, 1998.

[112] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Applied optics*, vol. 49, no. 36, pp. 6824–6833, 2010.

[113] M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[114] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Optics express*, vol. 17, no. 8, pp. 6368–6388, 2009.

[115] P. Song, X. Deng, , J. F. Mota, N. Deligiannis, P. L. Dragotti, and M. R. Rodrigues, "Multimodal image super-resolution via joint sparse representations induced by coupled dictionaries," *arXiv:1709.08680*, 2017.

[116] S. Friedland, Q. Li, and D. Schonfeld, "Compressive sensing of sparse tensors," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4438–4447, 2014.